

# Volatility dynamics for the S&P 500: Further evidence from non-affine, multi-factor jump diffusions

Andreas Kaeck<sup>a</sup>, and Carol Alexander<sup>b,\*</sup>

<sup>a</sup>*Swiss Institute of Banking and Finance, University of St. Gallen, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland.*

<sup>b</sup>*ICMA Centre, Henley Business School at the University of Reading, Reading, RG6 6BA, UK.*

This Version: 5 July 2012

---

## Abstract

We apply Markov chain Monte Carlo methods to time series data on S&P 500 index returns, and to its option prices via a term structure of VIX indices, to estimate eighteen different affine and non-affine stochastic volatility models with one or two variance factors, and where jumps are allowed in both the price and the instantaneous volatility. The in-sample fit to the VIX term structure shows that the second (stochastic long-term volatility) factor is required to fit the VIX term structure. Out-of-sample tests on the fit to individual option prices, as well as in-sample tests, show that the inclusion of jumps is less important than allowing for non-affine dynamics. The estimation and testing periods together cover more than twenty-one years of daily data.

*JEL classification:* G13; C13; C63

*Keywords:* Gibbs sampler; Instantaneous volatility dynamics; MCMC; Particle filter; S&P 500 options; VIX

---

\*Corresponding author. Tel.: +44 118 378 8239; fax: +44 118 931 4741.

E-mail addresses: [andreas.kaeck@unisg.ch](mailto:andreas.kaeck@unisg.ch) (A. Kaeck), [c.alexander@icmacentre.ac.uk](mailto:c.alexander@icmacentre.ac.uk) (C. Alexander).

We are grateful to an anonymous referee for insightful and constructive comments that helped to improve the paper considerably.

## 1. Introduction

The study of models for instantaneous volatility dynamics is important for many reasons. Accuracy of such a model improves option-hedging performance and the pricing of complex contingent claims. It also provides a better prediction for the realized variance that determines the gains and losses from trading variance swaps, and the realized characteristics that determine the value of associated products such as skewness swaps. Moreover, these models have the potential to unify the pricing of equity and volatility derivatives. This is especially important because trading in volatility derivatives (variance swaps, futures and options on the VIX index and exchange-traded notes based on constant-maturity VIX futures) has increased enormously during the last few years.

Early papers on the price dynamics of major equity indices focus on improving the affine stochastic volatility model of Heston (1993). Three main extensions have emerged. First, Bates (1996) and Bakshi *et al.* (1997) augment the Heston model with price jumps to account for the possibility of discontinuities in the price process. The Heston stochastic volatility diffusion seems unable to explain large return outliers, such as during the October 1987 crash. Hence, price jumps have become an important feature in many option pricing models. Second, Bates (2000) proposes a second diffusive variance factor, a specification that has the potential to alleviate the unrealistic volatility term-structure implied by single-factor variance models.<sup>1</sup> And third, some authors propose further jumps in the variance process of the Heston model. Motivated by findings in the time-series literature (most prominently Eraker *et al.*, 2003), Eraker (2004) and Broadie *et al.* (2007) investigate the role of variance jumps for pricing S&P 500 index and futures option contracts, respectively. Jumps in the instantaneous variance process should result in rare but dramatic changes in the variance evolution, and this feature might be necessary to capture the large changes in derivatives prices that occur during periods of market distress.

Recently, non-affine stochastic volatility models, such as GARCH or general CEV-type diffusions, have been found to capture the dynamics of major equity indices much better than the affine

---

<sup>1</sup>See also Christoffersen *et al.* (2009). More recently, Gruber *et al.* (2010) extend this model and allow the variance to be driven by correlated variance factors.

specification used in Heston (1993). For instance, the findings in Jones (2003) indicate the need for non-affine variance dynamics and Chernov *et al.* (2003) show that relaxing the affine constraint can reduce the need to add jumps. Christoffersen *et al.* (2010) find strong support for continuous-time GARCH specifications not only for spot but also for S&P 500 index option data. To date, however, most of the evidence regarding the superiority of non-affine variance models is presented in the time-series literature (Ignatieva *et al.*, 2009 and Kaeck and Alexander, 2011). Non-affine specifications lack closed-form solutions to vanilla option prices so there is much less work on the specification of non-affine volatility processes in the option pricing literature.

The main purpose of this study is to investigate extensions of non-affine variance dynamics and present comprehensive empirical results regarding their option pricing performance. Departing from the affine-only literature, we consider a general CEV-type stochastic volatility model augmented by jumps in both the price and variance process and by a second stochastic factor which drives the long-term variance level. This model nests many affine and non-affine specifications that have previously been considered and also many that have not. Thus, we contribute to the empirical option pricing literature in two important ways: we examine the need for price and/or variance jumps within non-affine as well as affine variance specifications; and we investigate the specification of long-term variance dynamics. The second contribution is important because the non-affine models that have previously been considered generate unrealistic volatility term structures. Hence, multi-factor extensions to non-affine specifications are potentially as important as they are for affine models. To the best of our knowledge we are the first to study such models using both spot and option time series data on the S&P 500 index.

The general model presented in this paper provides a unified framework for studying virtually all the combinations of jumps and diffusions that have been considered in the literature to date. Affine models are much easier to use so it is important to determine whether affine models with two stochastic volatility factors can produce a similar performance to one-factor non-affine models; or, for instance, whether a jump in the variance process of an affine stochastic volatility model can mitigate the need for non-affine dynamics. We further ask whether it is better to add jumps to

a model, or to increase the number of factors; and whether non-affine as well as affine models can be improved at all by adding jumps to the price and/or variance processes. Answers to these questions would resolve many questions that remain outstanding from the current literature. The existing empirical results regarding the dynamics of instantaneous variance are not conclusive. Some research highlights the importance of non-linear diffusions, other papers focus on the role of jumps or multiple factors. Some studies restrict their analysis to an affine model structure, while other papers examine non-linearities using non-affine models. The benefit of our framework is that it provides a general model within which all of these features can be compared.

Vanilla option prices are highly informative about stochastic volatility dynamics, but calibration to option prices alone ignores a second important source of information, namely the spot dynamics of the option underlying. For this reason, there is a large and growing literature on the use of both option price and underlying time-series data for the model calibration (see, for instance, Chernov and Ghysels, 2000 or Eraker, 2004). Most of this literature employs affine dynamics. The direct calibration of non-affine models to time series data is very difficult to handle because discrete-time transition probabilities or characteristic functions are generally unavailable in closed form. We use long time series on the S&P 500 index, and on its option prices as captured by a term structure of VIX volatility indices. This approach is particularly suited for our model specifications as the theoretical values for VIX indices can be derived for general CEV-type variance processes and empirical values of VIX indices can be constructed from S&P 500 index options. Our data covers more than twenty-one years of daily option data, with almost seven years reserved at the end for out-of-sample option pricing tests.

Since we fit our models to a term-structure of VIX indices our study relates to recent research on the variance-swap term structure, such as Amengual (2009) or Egloff *et al.* (2010). Both papers study two-factor variance specifications built from affine variance processes, whereas our focus is on non-affine model specifications and also on their out-of-sample option pricing performance. In addition, these papers estimate variance dynamics from the term structure of variance swaps

whereas we build VIX indices for different maturities.<sup>2</sup> Our estimation methodology is also related to the procedure in Lin (2007), Duan and Yeh (2010) and others. VIX indices are tractable, and they are also highly informative about variance dynamics, so studies that estimate variance dynamics from the VIX index, e.g. Zhang *et al.*, 2010 and Zhu and Lian, 2011, are becoming increasingly popular.

Our main findings can be summarized as follows. First, non-affine one-factor diffusion models clearly out-perform their affine counterparts both in-sample and out-of-sample. The out-of-sample root mean square error of affine specifications is approximately 20% higher than for non-affine models; this provides even stronger evidence in favor of non-affine specifications than the findings reported in Christoffersen *et al.* (2010). Second, the inclusion of jumps is less important than allowing for non-affine dynamics, indeed additional jumps only improve the fit of one-factor stochastic volatility models in-sample, not out-of-sample; the affine models class profits the most from additional jumps whereas jump-augmented non-affine models show no improvement over simple diffusion models in out-of-sample tests. Third, two-factor models are consistent with the observed term-structure of VIX indices, fitting both short-term and long-term prices well, whereas one-factor models are only adequate for capturing short-term price dynamics. The improvement of two-factor models over their single-factor counterparts becomes important for option maturities greater than six months. For these options we find that the out-of-sample root mean square error is more than 20% lower than for one-factor specifications. Consistent with earlier findings, non-affine two-factor models also significantly outperform their affine counterparts. We demonstrate that these results are robust to several distinct market regimes in our sample.

We proceed as follows: Section 2 introduces the affine and non-affine multi-factor jump diffusion models; Section 3 describes the data set and the VIX term structure; Section 4 describes our econometric estimation methodology; Section 5 describes the in-sample and out-of-sample specification tests; Section 6 presents all the empirical results, including estimation results for the

---

<sup>2</sup>As discussed later, in the presence of jumps there is a small difference between variance swaps and squared VIX indices.

eighteen possible models, in-sample fit and out-of-sample specification tests; Section 7 concludes.

## 2. Multi-factor jump diffusions

Under the risk-neutral probability measure  $Q$  we assume that the dynamics of the S&P 500 index, denoted  $s_t$ , are described by the following stochastic differential equations:

$$ds_t = (\mu^Q - \lambda\bar{\mu})s_t dt + \sqrt{v_t} s_t dw_t^{Q,s} + (e^{\xi_t^s} - 1) s_t dz_t, \quad (1)$$

$$dv_t = \kappa_v^Q(m_t - v_t) dt + \sigma_v (v_t)^{\gamma_v} dw_t^{Q,v} + \xi_t^v dz_t, \quad (2)$$

$$dm_t = \kappa_m^Q(\theta_m^Q - m_t) dt + \sigma_m (m_t)^{\gamma_m} dw_t^{Q,m}, \quad (3)$$

where  $\mu^Q = r - q$ , with  $r$  being the risk-free rate and  $q$  the dividend yield,  $v_t$  is the diffusive instantaneous variance process of  $s_t$  with volatility parameter  $\sigma_v$ ,  $\kappa_v^Q$  is its rate mean reversion to  $m_t$ , the long-term variance level with volatility parameter  $\sigma_m$ , and  $\kappa_m^Q$  is its rate mean reversion to the constant parameter  $\theta_m^Q$ . We assume  $dw_t^{Q,v} = \varrho dw_t^{Q,s} + \sqrt{1 - \varrho^2} dw_t^Q$ , with the three Brownian motion processes  $w_t^{Q,s}$ ,  $w_t^Q$  and  $w_t^{Q,m}$  being independent, so that  $dw_t^{Q,v}$  and  $dw_t^{Q,s}$  have correlation  $\varrho$ . We suppose that jumps occur only in the index and its instantaneous variance and that these jumps are contemporaneous, occurring at random times depending on increments of the Poisson process  $z_t$  with intensity  $\lambda$ . Following standard practice (see Eraker *et al.*, 2003, Eraker, 2004 or Broadie *et al.*, 2007) we assume the variance jump sizes are exponentially distributed with  $\xi_t^v \sim \mathcal{E}(\mu_v)$ , and return shocks are normally distributed with  $\xi_t^s \sim \mathcal{N}(\mu_s, \sigma_s)$ , so that  $\bar{\mu} = \exp\{\mu_s + \frac{1}{2}\sigma_s\} - 1$ .<sup>3</sup>

The general model (1) – (3) nests several well-known specifications. First, restricting  $m_t$  to be the constant  $\theta_v^Q$  yields the class of one-factor stochastic volatility (SV) models. Further setting  $\gamma_v = \frac{1}{2}$  we have an affine class of SV models that, following Heston (1993), has been extensively studied in the academic literature.<sup>4</sup> We shall study three models in this class, labeled Ji-A1 where

<sup>3</sup>We do not estimate a jump correlation parameter because Eraker *et al.* (2003) find no significance for such a correlation, attributing this finding to the difficulty in estimating the parameter precisely. For this reason, Broadie *et al.* (2007) exclude correlated jumps a priori.

<sup>4</sup>This model provides nearly closed-form transition densities for the two state variables and leads to tractable quasi-

$i = 0, 1, 2$  according as there are no jumps ( $z_t = 0$  for all  $t$ ), jumps in the price alone ( $\mu_v = 0$ ), and contemporaneous price and volatility jumps, respectively. If instead of  $\gamma_v = \frac{1}{2}$  we set  $\gamma_v = 1$  we obtain the one-factor GARCH SV model family which, following Nelson (1990), has also been studied by Christoffersen *et al.* (2010) and Kaeck and Alexander (2011). These models are labeled  $Ji$ -G1 with  $i$  being as previously defined. Finally, treating  $\gamma_v$  as a freely estimated parameter gives the constant elasticity of variance (CEV) class of SV models, labeled  $Ji$ -C1, which have been studied by Chacko and Viceira (2003), Ait-Sahalia and Kimmel (2007) and others. The corresponding two-factor SV models, which include a stochastic long-term variance diffusion, are denoted  $Ji$ -A2,  $Ji$ -G2 and  $Ji$ -C2 respectively. Thus the general model (1) – (3) nests a total of eighteen different specific models to be estimated.

We shall calibrate the model parameters using time series of market prices of the S&P 500, and a term structure of VIX indices. Thus we specify the transition between the risk-neutral measure,  $Q$  and the real-world probability measure,  $P$ . Often the dynamics under  $P$  are derived from a candidate stochastic discount factor which yields a simple and tractable transition between the two measures. For instance, the change of measure in Duffie *et al.* (2000) guarantees that both real-world and risk-neutral processes are affine. Theoretically more general risk premium specifications, as discussed in Cheridito *et al.* (2007), are possible. We use similar transformations to those applied in Christoffersen *et al.* (2010) and Egloff *et al.* (2010), hence the change of measure used in this paper includes two variance risk premia,  $\eta_v \equiv \kappa_v^Q - \kappa_v^P$  and  $\eta_m \equiv \kappa_m^Q - \kappa_m^P$  in addition to the price risk premium  $\eta_s \equiv \mu^Q - \mu^P$ . Thus, we specify the dynamics under the real-world measure  $P$  as

$$ds_t = (\mu^P - \lambda\bar{\mu})s_t dt + \sqrt{v_t}s_t dw_t^{P,s} + (e^{\xi_t^s} - 1)s_t dz_t, \quad (4)$$

$$dv_t = (\kappa_v^Q m_t - \kappa_v^P v_t) dt + \sigma_v(v_t)^{\gamma_v} dw_t^{P,v} + \xi_t^v dz_t, \quad (5)$$

$$dm_t = \kappa_m^P(\theta_m^P - m_t) dt + \sigma_m(m_t)^{\gamma_m} dw_t^{P,m}, \quad (6)$$

---

analytic vanilla option pricing formulae. See Heston (1993), Bates (1996), Bakshi *et al.* (1997), Duffie *et al.* (2000), Pan (2002) and many others.

where  $\theta_m^P = \kappa_m^Q \theta_m^Q / \kappa_m^P$ . We again assume the variance jump sizes are exponentially distributed with  $\xi_t^v \sim \mathcal{E}(\mu_v)$ , and return shocks are normally distributed with  $\xi_t^s \sim \mathcal{N}(\mu_s, \sigma_s)$ . We shall discuss deviations from this zero jump-risk premium assumption below.

### 3. Data and VIX term structure

For the calibration of the competing models we use daily close price data for the spot S&P 500 index and build VIX indices for different maturities. The VIX index is a popular measure of the implied volatility of S&P 500 index options traded at the Chicago Board Options Exchange (CBOE). It is central to our work because the squared VIX is a model-free estimate of the average integrated 30-day variance for all the models we consider, with only a small adjustment in the presence of jumps. Indeed, following Lin (2007) and others, the theoretical value of the VIX index of maturity  $T_i$  at time  $t$ , for  $i = 1, \dots, N_t$  may be written

$$vix_t(v_t, m_t, T_i | \Theta)^2 \times 10^{-4} = 2\lambda \left\{ e^{\mu_s + \frac{1}{2}(\sigma_s)^2} - 1 - \mu_s \right\} + \frac{1}{T_i - t} E_t^Q \left[ \int_t^{T_i} v_s ds \right], \quad (7)$$

where  $\Theta$  collects all the (real-world and risk-neutral) parameters of the model,  $T_i$  denotes the maturity date of the options used for the VIX calculation and  $E_t^Q[\cdot]$  is the conditional expectation at time  $t$  under the measure  $Q$ . In our framework, VIX indices for maturities converging to zero and infinity can be used to identify the instantaneous variance  $v_t$  and the long-term variance level  $m_t$ . In addition, the integrated variance can be derived even for non-affine models, which is a huge advantage in our framework.

Lin (2007), Duan and Yeh (2010) and others link the dynamics of the theoretical value (7) for a VIX index to its observable value,  $vix_t^*(T_i)$ , which is computed directly from vanilla option prices as described below. This way, they use CBOEs 30-day VIX index to calibrate both real-world and risk-neutral parameters of an option-pricing model. Extending this approach, we use a term-structure of VIX indices to calibrate both the risk-neutral parameters in (1) – (3) and the



real-world parameters in (4) – (6) simultaneously, using three time-series: of the S&P 500, and two time series for VIX indices of different maturities. One short and one long-term VIX index is sufficient to identify the two state-variables of our models, as discussed below.

If there are no jumps in the price process, then the VIX index of maturity  $T_i$  coincides with a continuously-monitored variance swap with the same maturity, and data on such swap rates could be used for calibration as an alternative to VIX indices. In the presence of jumps there is a small difference between the two quantities, see Broadie and Jain (2008) and Duan and Yeh (2010). Nevertheless, for reasonable parameter values this difference is small and our calibration approach can be regarded as almost identical to those that employ variance swap rates, such as Amengual (2009) and Egloff *et al.* (2010). The advantage of the VIX term-structure is that it can be constructed from option prices directly even in the presence of jumps, as shown by Carr and Wu (2006). This allows us to build a data set that covers more than two decades of daily data. Broadie *et al.* (2007) point out that when dealing with jump models it is essential to calibrate to a long data set which encompasses as many diverse data periods as possible in order to pin down the jump behavior of the S&P 500 index dynamics.

The VIX construction methodology used by the CBOE (see CBOE, 2009) has been found to provide systemic biases – see Jiang and Tian (2007) for a detailed discussion. Therefore, we deviate from the CBOE procedure in the following ways: (a) as in Carr and Wu (2009) we first translate the out-of-the-money option prices into implied volatilities and then use an interpolation-extrapolation scheme to control for both the discretization and the extrapolation error in the CBOE methodology. For the interpolation we use, as in Bliss and Panigirtzoglou (2004), a spline with a smoothing parameter of  $\lambda = 0.999$  and extrapolate the implied volatility curves linearly using the implied volatility of the last observed strike at both ends; (b) for short-term maturities we found that single quotes on deep out-of-the-money calls or puts can make the calculation of the VIX unstable and hence we discard option quotes from the tails of the implied volatility curves that are more than 0.05 in moneyness (defined as strike over futures value) away from the next quoted option. This significantly improves the stability of the calculations on a few days in the sample;

(c) we use the 3-month US LIBOR rate as a substitute for the risk-free interest rate as in Kostakis *et al.* (2011).

For the construction of VIX indices we use daily close S&P 500 index option data that covers more than two decades, from January 1990 until September 2011. To date this is one of the most extensive option data sets used in the literature. Our data consists of the best bid and offer quote at the market close and the identifying characteristics for each option (option type, strike and maturity). Altogether the options data set consists of more than 3.6 million quotes. We further use a range of standard filters to remove illiquid or erroneous quotes from the data set. First, we only consider options with more than one week to expiration. This is to ensure that our results are not biased by illiquidity and market-microstructure issues that affect options with very short time to expiration. Second, as in Gruber *et al.* (2010) we remove options for which the mid quote (simple average of bid and ask) is below 0.375\$, to avoid effects from minimum tick sizes. We also check for no-arbitrage conditions as in Bakshi *et al.* (1997). We then construct VIX indices with maturities of 30 and 360 days. S&P 500 option data are also available for maturities of more than one year but they are not very frequently traded, especially at the beginning of our sample when the number of quotes is substantially lower than in recent years. Nevertheless, we do keep the quotes on options with maturities longer than one year for the out-of-sample testing, since such options are potentially interesting to distinguish between one-factor and two-factor SV specifications.

Figure 1 depicts the daily S&P 500 index and its log returns over the sample period, as well as the 30-day and 360-day VIX indices. Table 1 provides descriptive statistics for the two VIX indices. The sample accommodates some very diverse sub-periods, including the dot-com bubble and its aftermath, the credit boom between 2004 and 2007, the very interesting period of the credit and banking crisis starting in 2007, and the sovereign debt crisis. Especially during the onset of the banking crisis in September 2008 we observe some very large outliers in the S&P 500 index returns.

#### 4. Econometric methodology

In order to estimate the risk-neutral and real-world parameters as well as to filter the unobservable state-variables for out-of-sample tests, we use a state-space representation. Following Eraker (2004), we use the Euler discretization of the model, employing a time step of one day, since the simulation studies in Eraker *et al.* (2003) and Li *et al.* (2008) demonstrate that the discretization bias of the Euler scheme is negligible for daily data. Thus, we obtain the discretized processes

$$\begin{aligned}\log s_t - \log s_{t-h} &= (\mu^P - \lambda\bar{\mu} - \frac{1}{2}v_{t-h})h + \sqrt{v_{t-h}} \varepsilon_t^{P,s} + (e^{\xi_t^s} - 1) b_t, \\ v_t &= v_{t-h} + (\kappa_v^Q m_{t-h} - \kappa_v^P v_{t-h}) h + \sigma_v (v_{t-h})^\gamma \varepsilon_t^{P,v} + \xi_t^v b_t, \\ m_t &= m_{t-h} + \kappa_m^P (\theta_m^P - m_{t-h}) h + \sigma_m (m_{t-h})^\gamma \varepsilon_t^{P,m},\end{aligned}$$

where  $b_t$  is a Bernoulli process with parameter  $\lambda h$  and  $\varepsilon_t^{P,s}$ ,  $\varepsilon_t^{P,v}$  and  $\varepsilon_t^{P,m}$  are normal variates with zero mean and variance  $h$ . The innovations  $\varepsilon_t^{P,s}$  and  $\varepsilon_t^{P,v}$  have a correlation of  $\varrho$ , all other random variables are independent. The time-step used throughout this paper is  $h = 1/252$ , i.e. one trading day.

Because of the jumps in state variables standard Kalman filtering techniques (used, for instance, by Egloff *et al.*, 2010) will not apply to our models. Alternative estimation methodologies often rely heavily on simulation methods, for instance efficient method of moments (Gallant and Tauchen, 1996) or particle filters (see Pitt and Shephard, 1999 or Johannes *et al.*, 2009). In this paper we address the estimation with Bayesian estimation procedures, and in particular we use a Markov chain Monte Carlo (MCMC) sampler. MCMC methods for discrete-time SV models are introduced by Jacquier *et al.* (1994) and Eraker *et al.*, 2003 and Li *et al.*, 2008 show that these methods yield very accurate estimates for jump-diffusion processes. They are also particularly useful in our setup because their hierarchical structure simplifies the estimation of complex models, in particular models with several unobserved state variables such as jumps and stochastic volatility factors.

The center of interest in Bayesian statistics is the joint distribution of parameters and latent variables conditional on the observed data. This distribution may be defined in terms of the posterior density

$$\begin{aligned}
 p(\Theta, \mathbf{v}, \mathbf{m}, \mathbf{z}, \boldsymbol{\xi}^v, \boldsymbol{\xi}^s \mid \mathbf{s}, \mathbf{vix}^*(T_i)) &\propto p(\mathbf{s}, \mathbf{vix}^*(T_i) \mid \Theta, \mathbf{v}, \mathbf{m}, \mathbf{z}, \boldsymbol{\xi}^v, \boldsymbol{\xi}^s) \\
 &\times p(\mathbf{v}, \mathbf{m} \mid \boldsymbol{\xi}^v, \boldsymbol{\xi}^s, \mathbf{z}, \Theta) \times p(\boldsymbol{\xi}^v, \boldsymbol{\xi}^s \mid \mathbf{z}, \Theta) \times p(\mathbf{z} \mid \Theta) \times p(\Theta)
 \end{aligned} \tag{8}$$

where  $\mathbf{vix}^*(T_i)$  denotes the observable VIX index of maturity  $T_i$ , which is different from its theoretical value given by (7). Following Amengual (2009) and others, the time  $t$  value for the VIX index of maturity  $T_i$  that is computed from market prices of vanilla options on the S&P 500 and is assumed to be related to its theoretical value as

$$\mathbf{vix}_i^*(T_i) = \mathbf{vix}_t(v_t, m_t^Q, T_i \mid \Theta) \times e^{\varepsilon_{i,t}}, \tag{9}$$

where the theoretical value on the right hand side is computed using (7) and either (10) or (11), depending on whether the SV model has one or two factors. The observed value may deviate from the theoretical value due to discretization errors (Jiang and Tian, 2007), numerical errors in integration over option prices and/or in the interpolation and extrapolation of the implied volatility surface (Alexander and Leontsinis, 2011) and due to market microstructure effects such bid-ask spreads. The error term  $\varepsilon_{i,t}$ , described in more detail below, captures all these sources of noise in the measurement of the VIX indices.

The first density on the right of (8) is the likelihood of the observed data conditional on the model parameters and  $p(\Theta)$  is the prior for the structural parameters  $\Theta$ . The remaining bold vectors collect all state variables and the observed data, for instance  $\mathbf{v} = \{v_t\}_{t=0,h,\dots,T}$ . Knowing the posterior density we can obtain point estimates and standard errors of structural parameters, as well as the probability of jump events and jump size estimates for each day in our sample. Prior distributions are chosen such that they are relatively uninformative, hence our parameter estimates

are driven by the information in the data and not the prior. We choose prior distributions that are either identical to Eraker *et al.* (2003) or Amengual (2009) or use similarly uninformative distributions.

The Euler approximation of the continuous-time model in the real-world measure allows for a closed-form posterior density. To recover this posterior density we apply the standard Gibbs sampler (Geman and Geman, 1984). This approach achieves the goal of simulating from a multi-dimensional posterior distribution by iteratively drawing from lower-dimensional, so-called ‘complete’ conditional distributions. Repeated simulation of the posterior allows one to estimate all quantities of interest, such as posterior means and standard deviations for structural parameters and latent state variables. The Gibbs sampler forms a Markov chain whose limiting distribution is the posterior density, under mild regularity conditions. We draw each of the univariate complete conditional distributions by simulating them directly whenever possible (see for example the appendix of Li *et al.* (2008) for detailed derivations) and whenever this is not feasible we apply the ARMS Metropolis algorithm – see Gilks *et al.* (1995) or Li *et al.* (2008) for an application in a similar framework. This algorithm, despite being computationally intensive, allows us to generate very high acceptance rates in the Metropolis steps. Hence, the algorithm we apply is as close as possible to that employed by Eraker *et al.* (2003).

For the out-of-sample performance assessment we need to filter out the unobservable state variables. To fix ideas let  $\mathbf{L} := \{v, m, z, \xi^v, \xi^y\}$  denote the latent state variables and  $\mathbf{Y} := \{s, \mathbf{vix}^*(T_i)\}$  denote the data, so that the posterior density may be written

$$p(\Theta, \mathbf{L} | \mathbf{Y}) \propto p(\mathbf{Y} | \Theta, \mathbf{L}) \times p(\mathbf{L} | \Theta) \times p(\Theta).$$

Ideally, this density has to be updated for every out-of-sample date, however brute-force methods such as rolling window estimations are too time-consuming to achieve this goal. We therefore follow standard practice, which is to fix the structural parameters from the in-sample estimation and then to run a small-scale Gibbs sampler to obtain draws from  $p(\mathbf{L}_{t'} | \mathbf{Y}_{t'})$  for each out-of-sample

date  $t'$ . This sequential updating procedure is called the particle filter – see Pitt and Shephard (1999) and Johannes *et al.* (2009) for more details.

## 5. Specification testing

We separate the sample into an estimation (in-sample) sub-period from January 1, 1990 to December 31, 2004 and an out-of-sample sub-period from January 1, 2005 to September 30, 2011. This allows both the estimation and testing of alternative models on sub-samples that encompass a variety of market regimes. We estimate models using daily log returns on the S&P 500 index and the two VIX indices with maturities 30 days and 360 days. For the out-of-sample tests we use the same two VIX indices to filter the unobserved state variables.

For testing different model specifications in-sample we employ the deviance information criterion (DIC) developed by Spiegelhalter *et al.* (2002). The DIC is a likelihood-based goodness-of-fit statistic with an in-built penalty for the number of model parameters, which can be regarded as a generalization of the AIC (Akaike) and the Schwartz (BIC) information criterion. Whilst neither the AIC nor the BIC apply to Bayesian model choice, the DIC is particularly useful in model selection problems where the parameters have been obtained using MCMC. For SV models with jumps the DIC has been studied in detail by Berg *et al.* (2004), who find that the ranking of models generated by the DIC (a lower value being indicative of a better model) is comparable to the ranking generated by Bayes factors. Bayes factors are the most popular tool of model choice in Bayesian statistics but they are extremely difficult to apply to high-dimensional problems such as ours. The DIC, on the other hand, can be calculated from the output of the MCMC sampler and has been successfully used in similar frameworks, e.g. Ignatieva *et al.* (2009).

Another useful specification test is based on the model's ability to capture the term structure behavior of the VIX. For the one-factor SV specifications, with  $m_t = \theta_v^Q$ , the expected integrated

variance in (7) can be derived as (see Bollerslev and Zhou, 2002)

$$E_t^Q \left[ \int_t^{T_i} v_s \, ds \right] = \frac{1 - e^{-\kappa_v^Q (T_i - t)}}{\kappa_v^Q} v_t + \frac{\kappa_v^Q \theta_v^Q + \lambda \mu_v^Q}{\kappa_v^Q} \left\{ (T_i - t) - \frac{1 - e^{-(T_i - t) \kappa_v^Q}}{\kappa_v^Q} \right\}. \quad (10)$$

In the two-factor SV models, following Egloff *et al.* (2010), we obtain

$$\begin{aligned} E_t^Q \left[ \int_t^{T_i} v_s \, ds \right] &= \frac{1 - e^{-\kappa_v^Q (T_i - t)}}{\kappa_v^Q} \times v_t + \frac{\left(1 - e^{-\kappa_v^Q (T_i - t)}\right) \kappa_m^Q - \left(1 - e^{-\kappa_m^Q (T_i - t)}\right) \kappa_v^Q}{\kappa_m^Q (\kappa_m^Q - \kappa_v^Q)} \times m_t^Q \\ &+ \left\{ \theta_m^Q + \frac{\lambda \mu_v^Q}{\kappa_v^Q} \right\} \left\{ (T_i - t) - \frac{1 - e^{-\kappa_v^Q (T_i - t)}}{\kappa_v^Q} \right\} \\ &- \theta_m^Q \left\{ \frac{\left(1 - e^{-\kappa_v^Q (T_i - t)}\right) \kappa_m^Q - \left(1 - e^{-\kappa_m^Q (T_i - t)}\right) \kappa_v^Q}{\kappa_m^Q (\kappa_m^Q - \kappa_v^Q)} \right\}. \end{aligned} \quad (11)$$

The proximity between the VIX theoretical values based on a calibrated model and the observed values of the VIX given by (9) is an important model selection criterion. As in Eraker (2004), Amengual (2009) and others, we suppose  $\varepsilon_{i,t}$  follows a continuous-time Ornstein-Uhlenbeck process, so that the discrete observations on the VIX term structure errors follow a first order autoregressive, i.e. AR(1) process, with

$$\varepsilon_{i,t} = \rho_{\varepsilon_{i,t}} \varepsilon_{i,t-1} + \sigma_{\varepsilon_{i,t}} \psi_{i,t}, \quad \psi_{i,t} \sim \text{NID}(0, 1), \quad (12)$$

where the autocorrelation parameter  $\rho_{\varepsilon_{i,t}}$  should be small. Testing for such autocorrelation within the estimation sample provides a good insight to the ability of the models to capture the maturity behavior of vanilla options on the S&P 500, as reflected by the term structure of the VIX.<sup>5</sup>

For the out-of-sample performance testing we follow Bakshi *et al.* (1997), Eraker (2004), Christoffersen *et al.* (2010) and many others, by generating the difference between observed and theoretical vanilla option prices produced by each model, at strikes and maturities for which liquid

<sup>5</sup>The most common alternative to the error specification used in this paper is to assume iid errors such as in Egloff *et al.* (2010). We adopt a more general error specification because this allows us to capture potential misspecifications in the error term.

market data are available.<sup>6</sup> We use the market data to compute a weighted pricing error for each of the eighteen models considered. The pricing error on the  $i$ th option at time  $t$  is defined as

$$\tau_{i,t} = w_i [O_t(T_i, K_i, \alpha_i | \mathcal{M}) - O_t^{ma}(T_i, K_i, \alpha_i)], \quad (13)$$

where  $T_i$  and  $K_i$  denote the maturity and strike of the  $i$ th option,  $\alpha_i$  indicates the option type (put or call),  $O_t^{ma}(T_i, K_i, \alpha_i)$  is the market price at time  $t$ ,  $\mathcal{M}$  denotes the model and  $O_t(T_i, K_i, \Omega_i | \mathcal{M})$  is the model price at time  $t$ , and  $w_i$  is a weight. We follow Christoffersen *et al.* (2010) and others by setting  $w_i$  to the inverse Black-Scholes vega of the option. This allows one to interpret the pricing error as an approximation to an implied volatility error. To obtain the estimate for the model-based option price in (13) we use the standard Bayesian approach, which is to integrate out the uncertainty in the state variables, i.e.

$$O_t(T_i, K_i, \alpha_i | \mathcal{M}) = \int O_t(T_i, K_i, \alpha_i; \Theta, v_t, m_t | \mathcal{M}) dP(v_t, m_t | s_t, \text{vix}_t^*(T_i)).$$

Since most models in our study lack closed-form pricing formulae we estimate each model price by Monte-Carlo simulation with 100,000 simulated paths and ten discretization steps per day. Finally, to assess any bias and inefficiency in the model's theoretical option prices, we examine  $\sum \tau_{i,t}^2$ . Here the sum is taken over the entire out-of-sample period: first over all options, and then over some subsets of options as defined by moneyness or maturity. Clearly, the smaller the squared errors, the better the model's option prices fit to the observed prices.

## 6. Empirical results

This section reports the estimation results for both one-factor SV and two-factor SV specifications introduced in Section 2, and details on the results of the in-sample and out-of-sample tests outlined

---

<sup>6</sup>For this exercise we apply put-call-parity to obtain fair prices of S&P 500 futures underlying the options. It is standard in the literature to price the option on this futures price, as this obviates the need to estimate the unobservable dividend yield.



in Section 5.

### 6.1. Parameter estimation and goodness-of-fit

Table 2 reports the estimates for the structural parameters of the nine different one-factor SV specifications, and the DIC for each model. Our estimate for  $\gamma$  lies between the affine and the GARCH diffusion parameters, with values between 0.733 and 0.788. This is consistent with Ait-Sahalia and Kimmel (2007), although their estimate 0.65 is slightly lower than ours. The estimated correlation coefficients for our sample are between  $-0.754$  and  $-0.824$ , these values being in line with earlier research on the S&P 500 and VIX dynamics. Research focusing on time-series data on the S&P 500 report smaller values for this parameter, for instance Eraker *et al.* (2003) obtain a correlation estimate of only  $-0.36$  for the Heston model and up to  $-0.50$  in their jump specifications. Our estimates are relatively stable across different model specifications because the VIX indices and their correlations with the S&P 500 index returns provide a strong signal for this parameter.

An important difference between our results and those previously reported in the empirical VIX and variance swap literature is that we find only little support for negative variance risk premia. Although the sign of the risk premium in seven of the models is negative, it is small in comparison with the standard errors on the two mean reversion coefficient estimates, so there is no statistical significance for most of the variance risk premium estimates. This result seems at odds with the existing literature which uses the VIX or variance swaps in the estimation procedure. For instance, for the Heston model (J0-A1) Egloff *et al.* (2010) estimate  $\kappa_v^P$  to be 4.49 whereas our point estimate is 4.182, but with a standard error of 0.355. However, they find a large negative variance risk premium based on an estimated  $\kappa_v^Q$  of only 0.15, whereas our point estimate of  $\kappa_v^Q$  is 4.052 with a standard error of only 0.178. Similarly, using the standard 30-day VIX index, Zhu and Lian (2011) estimate  $\kappa_v^P$  to be 2.27, but under  $Q$  they estimate a near unit-root variance process with a  $\kappa_v^Q$  estimate of 0.25. The results of Duan and Yeh (2010) even imply explosive behavior of

the variance process under the risk-neutral measure.

How economically plausible are our estimates with respect to these earlier findings? To address this question it is instructive to examine the long-term variance levels under each probability measure, based on the variance risk premium estimates. For our parameter estimates, there is little difference between risk-neutral and real-world long-term variance: both are roughly 20%, quoted as an annual volatility. Note that both long-term and short-term VIX indices exhibit average levels of around 20% for our estimation sample period and consequently our mean-reversion parameter estimates and their implied variance risk premium appear to be economically reasonable. By contrast, the variance risk premium estimates of Egloff *et al.* (2010) imply a long-term volatility level of 34% under the risk-neutral measure but only 6% under the real-world probability measure: both these estimates are more extreme, especially the very low real-world long-term volatility level of 6%. Similarly, the results in Zhu and Lian (2011) imply a very high long-term volatility level under the risk-neutral measure of 63%.<sup>7</sup>

It is also likely that previously reported estimates are sensitive to the sample data. Indeed, most papers that report variance risk premium estimates not only use different models, they also differ significantly in terms of the data employed: in its frequency and sample period as well as using a variance term structure or just a single point. Broadie *et al.* (2007) claim that when the variance term structure is relatively flat, risk premia are very difficult to identify. Of course, one also needs to be very cautious when interpreting results from a potentially mis-specified model, such as the Heston model, but the results from all our models imply relatively small and economically insignificant risk premium estimates.

We now discuss the role of jumps. Allowing the intensity of the jumps to be different under  $P$  and  $Q$ , we first estimate all models with an additional jump risk premium. Finding that this addition makes no significant difference, except in one model, and that the qualitative and quantitative

---

<sup>7</sup>To check whether our methodology would yield results that are in line with these findings, we also estimate the affine one-factor model using only the 30-day VIX index. In this case we do find a significantly negative variance risk premium. Therefore, the reported differences between our findings and the literature can be attributed to the information in the variance term structure.

results remain unchanged for all the models, these results are not reported. Generally, the VIX appears to be rather uninformative regarding the jump size parameters. Indeed, the analytical form of the VIX pricing formulae implies that such parameters are difficult to identify, for instance the complete conditional distribution for  $\mu_s$  under the risk-neutral measure is bi-modal with one mode covering the negative and the other covering positive jumps. This is identical to the problem for variance swaps where, in the presence of price jumps, the variance swap rate depends only on  $\mu_s^2$  and hence the sign of the risk-neutral jump is not identified. Intuitively, only the size of jumps matters for their contribution to overall variance, not their sign. For this reason, estimating jump parameters from the VIX index or variance swaps is extremely difficult. Amengual (2009) appears to estimate jump parameters from the time-series of the S&P 500 only and Sepp (2008) restricts jumps to occur only in the variance process. Since we find no significant differences between risk-neutral and real-world jump intensities we use the information in the S&P 500 index only.

Our jump parameter estimates are roughly in line with those obtained by Eraker *et al.* (2003), Chernov *et al.* (2003) and Andersen *et al.* (2002). We estimate an expected frequency of between five and six jumps per year, which is similar to the estimate of Andersen *et al.* (2002). Although this is slightly larger than the jump frequency estimates of Eraker *et al.* (2003) and Chernov *et al.* (2003), the small difference may be explained by differences in estimation sample periods and estimation methodology. Other literature, based on calibration to option prices or variance swaps, is more inconclusive about the role of jumps. Here the expected frequencies range between less than one per year in Bakshi *et al.* (1997) and Zhu and Lian (2011) to more than 100 jumps per year in Duan and Yeh (2010).<sup>8</sup>

The in-sample fit statistics shown in Table 2 provide clear evidence regarding the role of the CEV parameter and the importance of jumps. Jumps improve the in-sample fit of all the models, even after adjusting for the increase in the number of parameters. The differences in the DIC values are relatively large.<sup>9</sup> For all three model classes variance jump models perform best, followed by

<sup>8</sup>Bakshi *et al.* (1997) estimate only 0.59 jumps per year using S&P 500 index options, and Zhu and Lian (2011) estimate 0.50 jumps per year using the joint time-series of S&P 500 and the 30-day VIX.

<sup>9</sup>Spiegelhalter *et al.* (2002) suggest that a difference of only 10 is significant, so our results are very strong.

the price jump specification and the pure diffusion model. The best performing model class is the GARCH specification followed by CEV specifications and the affine models. Note that the simple GARCH diffusion model performs even better than the affine price jump model, a finding that is in line with the results from the time-series of the S&P 500, see Kaeck and Alexander (2011). We conclude that (a) the GARCH and CEV models significantly outperform their affine counterparts and (b) that jumps in prices and jumps in variance are both important features of the data-generating process in the one-factor SV model class.

Table 3 presents parameter estimates for the two-factor variance specifications. Initially we estimate a mean-reverting long-term variance processes under both measures but because the risk-neutral drift has very little economic impact on VIX values, even for very long time to expiration, the drift parameters for  $m_t$  under the risk-neutral measure are highly unstable.<sup>10</sup> The main reason for their instability is that the mean-reversion rate is very slow (e.g. Amengual (2009) estimates a parameter of  $\kappa_m^Q = 0.057$ ) so it is difficult to identify a long-term value to which  $m_t$  converges. Some previous research (for instance Luo and Zhang, 2010) imposes the same restriction a priori.

Since the 30-day VIX provides a strong signal for  $v_t$ , parameter estimates of the variance processes under the two-factor specifications are largely similar to those obtained for one-factor models, in Table 2. In particular, we again find only weak evidence for a negative variance risk-premium. Although we estimate  $\kappa_v^P > \kappa_v^Q$  in all models, high standard errors on these parameters make their difference insignificant. Jump parameter estimates are also roughly in line with those in Table 2, although we find a higher number of jumps (ten to eleven per year) in the double-jump models. The CEV parameter for the variance process also increases to between 0.864 (for the double-jump model) to 0.960 (for the model with no jumps). As expected, the parameter estimates for the long-term variance imply a very slowly mean-reverting process under the real-world probability measure. We estimate  $\kappa_m^P$  between 0.16 and 0.43, so the characteristic time to

<sup>10</sup>For instance, readers can verify that there is negligible difference between the term structure curves generated by  $\kappa_m^Q = 0$  and  $\kappa_m^Q = 0.6$ , fixing the other parameters to reasonable values. For instance, we have set  $\mu_s^Q = -0.03$ ,  $\sigma_s^Q = 0.02$ ,  $\mu_v^Q = 0.01$ ,  $\kappa_v^Q = 4$ ,  $\lambda^Q = 5$  and  $\theta_m^Q = 0.02$  and  $m_t^Q = 0.025$ , and results for these or other parameters are available on request.

mean revert is between 2 and 5 years.

The in-sample fit statistics are again highly consistent across the three model classes. Moreover, despite the additional parameters, the DIC values in Table 3 are much lower than their counterparts in Table 2. This provides the first clear evidence that a stochastic long-term variance is a significant factor determining the S&P 500 index dynamics. Identical to the one-factor models, we find the lowest DIC values for the GARCH model class. The second best performing model class is – also consistent with the results for one-factor specifications – the CEV. Hence, the fit of non-affine models to the S&P 500 and VIX data is clearly better than the fit of the affine models. Interestingly, the need for jumps becomes questionable, now that we include a second variance factor. Indeed, our results provide evidence in favour of the most parsimonious, diffusion-only two-factor SV specifications. The addition of a second variance factor significantly reduces the need for a jump component and two-factor models augmented with jumps appear to be over-parameterized.

## 6.2. *Fitting the VIX term structure*

The estimated coefficients on the AR(1) process (12) for the VIX term structure errors are presented in Table 4. The upper part of this table contains the estimates corresponding to one-factor SV models. Here the estimates of the AR(1) parameters are very close to each other in all nine one-factor specifications, so they provide little information regarding the relative performance of alternative models. As hoped, the mean-reversion parameter  $\rho_\varepsilon$  is close to zero for the short-term index, with small positive values indicating that there is only weak autocorrelation in the error terms. By contrast, for the long-term VIX, the estimated parameters for  $\rho_\varepsilon$  are close to one and this indicates strong evidence against all nine one-factor SV specifications. This finding is unsurprising because one-factor SV models predict that long-term VIX indices converge relatively quickly to the square root of the constant long-term variance and the error process obviously picks up this mis-specification with near-unit-root behavior. Also note that the process parameters for the error of the 360 day VIX are almost identical in all model specifications because all models require

similar error processes to explain the difference between market values and the predicted (almost) constant model values.

The fit to the error dynamics (12) for the two-factor SV models is shown in the lower part of Table 4. The dynamics for the short-term VIX process is comparable to the fit obtained from the one-factor specifications. There is, however, a marked difference for the long-term variance: for diffusion and price-jump models the error between the log observed VIX and the log of the two-factor theoretical VIX now has an autocorrelation coefficient that is insignificantly different from zero. This implies that the addition of a second SV factor yields a model that is able to produce long-term VIX values that are in line with the observed data. As with our earlier results, the additional variance jump deteriorates the fit and yields higher autocorrelation coefficients.<sup>11</sup>

Overall, these results raise the question of whether a two-factor model is also consistent with VIX indices of intermediate maturities. The results in Amengual (2009) suggest that pricing errors for intermediate maturities are still relatively low compared with the strong misspecification in one-factor models. In order not to bias the likelihood function heavily against S&P 500 index dynamics, we have opted for as many VIX indices as variance factors. Examining the fit to more VIX indices and utilizing more variance factors, as in Gruber *et al.* (2010), is a potentially interesting avenue for future research.

### 6.3. Out-of-sample tests on option prices

We now hold the model parameters constant, at the values shown in Tables 2 and 3, and derive vanilla option prices for every day in the out-of-sample period. This period spans almost seven years, from January 1, 2005 until September 30, 2011 and contains more than 700,000 option prices with moneyness ( $K/F$ ) between 0.8 and 1.1 and time to maturity between 7 and 550 days.

---

<sup>11</sup>The ability to fit the VIX term structure is also related to the goodness-of-fit of the AR(1) process (12) as measured by the error volatility  $\sigma_\epsilon$ . In the two-factor models with no jumps, or just one jump in the price process, this volatility is smaller than it is when the VIX term structure is fitted using their one-factor counterparts. However, for the double jump models the error volatility increases when the second stochastic factor is added. This observation ties in with our earlier results and provides further evidence that including a jump to instantaneous volatility adds too much variability to the process when the long term volatility is stochastic.

Since we employ a huge number of alternative pricing models and calculate out-of-sample pricing errors by Monte Carlo, we follow numerous other papers in the literature (e.g. Christoffersen *et al.*, 2010) by using Wednesday option price data only. Then calculating implied volatility root mean square error (RMSE) over all options and over the entire out-of-sample period yields the results displayed in Table 5. The RMSE measures the dispersion of the model prices about the observed market prices, so the lower the RMSE the more accurate the model prices.

The results in Table 5 indicate that within the one-factor pure diffusion class, the CEV specification provides the smallest RMSEs whereas the affine class produces the largest. The differences between the RMSE of affine and non-affine models are substantial, the GARCH diffusion model has an out-of-sample pricing error which is more than 18% less than for the affine diffusion model (the Heston specification). The CEV model performs even better than the GARCH diffusion and yields an improvement of more than 20% over the affine specification. In fact, the improvements that we find for non-affine models are even greater than those reported in Christoffersen *et al.* (2010). The addition of a second factor in the J0 class contributes little to their out-of-sample pricing ability. Nevertheless, the ranking of the models remain intact, with the two-factor CEV model outperforming the affine specification by almost 18%.

Further augmenting these models with jumps does little to improve their performance. Indeed, the inclusion of jumps in the instantaneous variance again only improves the performance for one-factor affine models. Our results demonstrate that the inclusion of jumps is not necessary to attain a good option-pricing performance with the GARCH or CEV specifications, and simple non-affine diffusion models out-perform even jump-augmented affine specifications. Thus, while jumps can help affine models, the choice of accurate variance dynamics seems to be far more important than including price or volatility jumps.

In most of the current option-pricing literature the parameters are chosen to minimize an option-pricing objective function. Moreover, the models are often re-calibrated on a frequent basis, rather than holding parameters constant over a period of nearly seven years. As a result, absolute out-of-sample pricing errors in the related literature can be lower than those reported in Table 5

– see Christoffersen *et al.* (2010), for instance. By calibrating to option prices only indirectly, through use of a time-series on the VIX term structure, and by using a fixed set of parameters that were estimated over a very long in-sample period, we have traded the out-of-sample fit of the models for internal consistency. Broadie *et al.* (2007) have made a similar trade-off, and we note that the size of errors reported there are comparable to our values, even though the design of their out-of-sample pricing test differs markedly from ours.

Compared with the one-factor models we have found little or no improvement in the pricing performance of the two-factor models. At first sight this result seems at odds with our in-sample results, where substantial improvements were observed after including a time-varying long-term variance. However, there is a simple explanation for this, in that most option pricing data sets are heavily biased towards very short-term options for which the second factor is not very important. For instance, our sample includes over 90,000 options with maturity less than three months, but less than 9,000 options with maturity longer than one year. For this reason it is important to stratify the results by option maturity, so that we may judge the pricing improvement of the two-factor models fairly.

To study which option categories drive our out-of-sample results, we first stratify the RMSE into four categories, according as the option's maturity is 7–90 days, 91–180 days, 181–365 days and 366–550 days. Table 6 presents the RMSE errors for each model, along with the number of option prices falling into each category. These results are particularly enlightening about the performance of one-factor SV models relative to two-factor SV models: the RMSE is uniformly lower for the two-factor models, provided the aggregation is over options with maturities longer than 180 days. By including such a factor the implied volatility RMSEs for options with maturity between six months and one year can be reduced, on average, by 14%, and for options with maturity greater than one year the second factor reduces the RMSE by more than 20%. On the other hand, for short-term options the one-factor models provide the best predictions, with an average RMSE about 4% lower than the corresponding RMSE when long-term volatility is allowed to be stochastic. Overall, this provides strong evidence that the importance of a second variance factor



becomes visible for options with maturity exceeding 6 months.

Within the one-factor SV class, the results also demonstrate that the affine models have the lowest RMSE for options with maturity greater than six months. Indeed, in general, most of the improvement from using a non-affine specification is for short-term options. But this is to be expected because the advantage of the CEV/GARCH models is that their variance process can move more flexibly in the short run, which creates more realistic short-term option prices and, as noted already for the two-factor model, may preclude the need to add jumps. Overall we can conclude that for short-term options the CEV exponent has a strong influence on out-of-sample results, whereas for longer-term options it is important to model a stochastic long-term variance factor. However, the dynamics of this long-term factor are less important, with affine and non-affine models providing similar RMSEs.

Table 7 presents the implied volatility errors aggregated by the option's moneyness category, with four different categories and the number of option prices in each category shown in the first column of the table. In the one-factor diffusion model class the CEV model outperforms the remaining two specifications in three out of four moneyness categories, with the GARCH performing marginally better only for options within the 1.03 to 1.1 moneyness band. The RMSE for at-the-money options is similar in all three model classes. For deep in/out-the-money calls/puts (moneyness 0.8 to 0.9) the inclusion of jumps only improves the performance of the Heston model. But when jumps are added to the non-affine one-factor models, the fit to out-of-the-money option prices deteriorates. This result is intuitive, because one would expect that the ability of the GARCH and CEV models to provide more flexible variance dynamics, without the need for jumps, would play the biggest role when pricing out-of-the money options. Within the two-factor models the GARCH and CEV specifications again outperform their affine counterparts, but now jumps have little effect on a model's performance, now even within the affine model class.

Our out-of-sample option data set covers some very distinct time periods, including the stable credit-boom years at the beginning of the sample, the banking crisis in the middle and the sovereign debt crisis towards the end of the sample. To understand if any of the out-of-sample results are

driven by a specific market regime, in particular to investigate the robustness of our findings on the superiority of GARCH and CEV models, we break up the sample into calendar months and aggregate monthly RMSE over various sub-samples. To concentrate on our main results, we provide detailed analyses only for one model in each model category. First, we choose J2-A1 since it provides the best in- and out-of-sample performance in the affine one-factor model class. For the other five categories, we use the simple diffusion models i.e. J0-A2, J0-G1, J0-G2, J0-C1, J0-C2 since they exhibit the best, or close to the best pricing performance. Our results would be even stronger if we also used J0-A1 instead of the contemporaneous jump model J2-A1; the choice of J2-A1 is to give the affine models the best chance of success while restricting the number of models to focus on the main conclusions.

For each month in the sample we first calculate the RMSE as in Table 5. Then, for each calendar month in our sample we calculate the average RMSE for the previous twelve months and roll forward monthly, to produce the graph depicted in Figure 2. This graph provides strong support for our previous findings. For the one-factor models, with rolling RMSEs depicted by the continuous lines, the CEV specification out-performs the affine representative in every month over the entire period, which provides striking evidence in favor of the standard CEV diffusion for instantaneous variance. Nevertheless, the GARCH model performs slightly better until the post-banking crisis period. By contrast, once we add the second SV factor, the CEV specification provides an improvement on GARCH only in the pre-banking crisis period.

The performance of the two-factor affine model was dismal during the onset of the banking crisis, in September 2008. Thus, even if we augment the affine class with jumps so that it can compete with the GARCH and CEV diffusions, the affine-jump model results can be poor during excessively volatile periods. Both CEV and GARCH two-factor specifications clearly out-performed their one-factor counterparts only during the banking and Eurozone crises. The reason for this is that volatility was bounded at around 20% prior to September 2008, which is also close to the estimate for the long-term volatility in the one-factor models. Thus, there was little need for a stochastic long-term variance factor prior to the banking crisis. But during the crisis, variance

rapidly increased, with the VIX reaching a high of more than 80% by October 2008. The one-factor models predict a quick mean reversion of the variance, while the two-factor models now provide substantially different long-term variance values. Since variance remained high for a long time after the onset of the crisis it seems plausible that the second variance factor drives much of the improved performance during and after the high-volatility regime.

## **7. Conclusion**

This study fills a notable gap in the current literature on the dynamics of instantaneous volatility. We augment non-affine as well as affine option pricing models with jumps and a second stochastic variance factor and provide specification tests on eighteen different models over a sample period of almost twenty-one years. Some of our empirical results confirm previous findings (e.g. a two-factor SV specification is required for fitting the term structure of VIX indices) but we also find several new results which should be important for end-users including option traders, variance swap traders and risk managers.

By focusing on the inclusion of jumps in a general, multi-factor stochastic variance model we show that it is important to adopt a two-factor model with a mean-reverting long-term variance diffusion. However, during tranquil markets, the one-factor SV models can fit option prices almost as well as two-factor SV models, provided the dynamics are non-affine. We also provide clear evidence that non-affine models out-perform their affine counterparts, and that once the diffusion dynamics are accurate there is little need to add jumps to the price and/or instantaneous variance. Although the affine specification is much easier to employ than non-affine models (given the closed form solutions that it offers) the problem with affine volatility dynamics is that they are insufficiently flexible to capture the rapid movements that characterize volatility. Even adding jumps to the affine framework produces a performance that is significantly inferior to the non-affine class without jumps, especially during volatile markets.

Our findings regarding the variance risk premium are mixed. On average, we estimate a neg-

ative premium but these estimates are statistically significant in only few model specifications. Probably this is because, as previously noted by Broadie *et al.* (2007), the average variance term structure is relatively flat.

## References

- Ait-Sahalia, Y. and Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83, 413–452.
- Alexander, C. and Leontsinis, S., 2011. Model risk in variance swap rates. Working Paper, University of Reading.
- Amengual, D., 2009. The term structure of variance risk premia. Working Paper, Princeton University.
- Andersen, T. G., Benzoni, L., and Lund, J., 2002. An empirical investigation of continuous-time equity return models. *Journal of Finance*, 57(3), 1239–1284.
- Bakshi, G., Cao, C., and Chen, Z., 1997. Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5), 2003–49.
- Bates, D. S., 1996. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. *Review of Financial Studies*, 9(1), 69–107.
- Bates, D. S., 2000. Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94(1-2), 181–238.
- Berg, A., Meyer, R., and Yu, J., 2004. Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics*, 22(1), 107–120.
- Bliss, R. R. and Panigirtzoglou, N., 2004. Option-implied risk aversion estimates. *Journal of Finance*, 59(1), 407–446.

- Bollerslev, T. and Zhou, H., 2002. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109(1), 33–65.
- Broadie, M. and Jain, A., 2008. The effect of jumps and discrete sampling on volatility and variance swaps. *International Journal of Theoretical and Applied Finance*, 11(08), 761–797.
- Broadie, M., Chernov, M., and Johannes, M., 2007. Model specification and risk premia: Evidence from futures options. *Journal of Finance*, 62(3), 1453–1490.
- Carr, P. and Wu, L., 2006. A tale of two indices. *Journal of Derivatives*, 13(3), 13 – 29.
- Carr, P. and Wu, L., 2009. Variance risk premiums. *Review of Financial Studies*, 22(3), 1311–1341.
- CBOE, 2009. The CBOE volatility index (R) - Vix (R). White Paper, Available from [www.CBOE.com](http://www.CBOE.com).
- Chacko, G. and Viceira, L. M., 2003. Spectral GMM estimation of continuous-time processes. *Journal of Econometrics*, 116(1-2), 259–292.
- Cheridito, P., Filipovic, D., and Kimmel, R. L., 2007. Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics*, 83, 123–170.
- Chernov, M. and 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(3), 407–458.
- Chernov, M., Gallant, A. R., Ghysels, E., and Tauchen, G., 2003. Alternative models for stock price dynamics. *Journal of Econometrics*, 116(1-2), 225–257.
- Christoffersen, P., Heston, S., and Jacobs, K., 2009. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 55(12), 1914–1932.

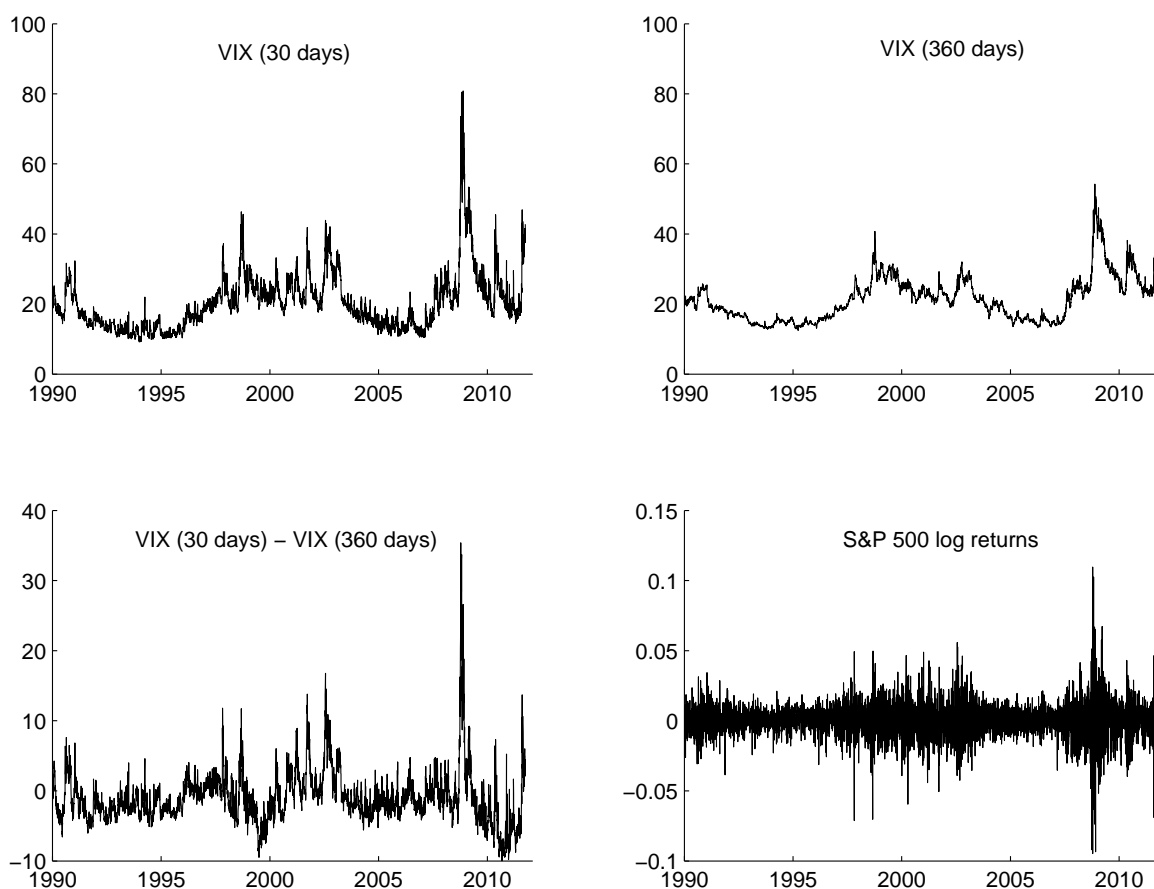
- Christoffersen, P., Jacobs, K., and Mimouni, K., 2010. Volatility dynamics for the S&P 500: Evidence from realized volatility, daily returns, and option prices. *Review of Financial Studies*, 23(8), 3141–3189.
- Duan, J.-C. and Yeh, C.-Y., 2010. Jump and volatility risk premiums implied by VIX. *Journal of Economic Dynamics and Control*, 34(11), 2232–2244.
- Duffie, D., Pan, J., and Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6), 1343–1376.
- Egloff, D., Leippold, M., and Wu, L., 2010. Variance risk dynamics, variance risk premia, and optimal variance swap investments. *Journal of Financial and Quantitative Analysis*, 45(5), 1279–1310.
- Eraker, B., 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance*, 59(3), 1367–1404.
- Eraker, B., Johannes, M., and Polson, N., 2003. The impact of jumps in volatility and returns. *Journal of Finance*, 53(3), 1269–1300.
- Gallant, A. R. and Tauchen, G., 1996. Which moments to match? *Econometric Theory*, 12(04), 657–681.
- Geman, S. and Geman, G., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian resoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Gilks, W. R., Best, N. G., and Tan, K. K., 1995. Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics*, 44(4), 455–472.
- Gruber, P., Tebaldi, C., and Trojani, F., 2010. Three make a smile - Dynamic volatility, skewness and term structure components in option valuation. Working Paper, University of Lugano.

- Heston, S. L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2), 327–43.
- Ignatieva, K., Rodrigues, P., and Seeger, N., 2009. Stochastic volatility and jumps: Exponentially affine yes or no? An empirical analysis of S&P500 dynamics. Working Paper, Goethe University Frankfurt.
- Jacquier, E., Polson, N. G., and Rossi, P. E., 1994. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 12(4), 371–89.
- Jiang, G. J. and Tian, Y. S., 2007. Extracting model-free volatility from option prices: An examination of the VIX index. *Journal of Derivatives*, 14(3), 35–60.
- Johannes, M. S., Polson, N. G., and Stroud, J. R., 2009. Optimal filtering of jump diffusions: Extracting latent states from asset prices. *Review of Financial Studies*, 22(7), 2759–2799.
- Jones, C. S., 2003. The dynamics of stochastic volatility: Evidence from underlying and options markets. *Journal of Econometrics*, 116, 181–224.
- Kaeck, A. and Alexander, C., 2011. Stochastic volatility jump-diffusions for european equity index dynamics. *European Financial Management*, forthcoming.
- Kostakis, A., Panigirtzoglou, N., and Skiadopoulos, G., 2011. Market timing with option-implied distributions: A forward-looking approach. *Management Science*, forthcoming.
- Li, H., Wells, M. T., and Yu, C. L., 2008. A Bayesian analysis of return dynamics with Levy jumps. *Review of Financial Studies*, 21(5), 2345–2378.
- Lin, Y., 2007. Pricing vix futures: Evidence from integrated physical and risk-neutral probability measures. *Journal of Futures Markets*, 27(12), 1175–1217.
- Luo, X. and Zhang, E., 2010. The term structure of VIX. Working Paper, University of Hong Kong.

- Nelson, D. B., 1990. ARCH models as diffusion approximations. *Journal of Econometrics*, 45(1-2), 7–38.
- Pan, J., 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63, 3–50.
- Pitt, M. K. and Shephard, N., 1999. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94, 590–599.
- Sepp, A., 2008. Pricing options on realized variance in the Heston model with jumps in returns and volatility. *Journal of Computational Finance*, 11(4), 33–70.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal Of The Royal Statistical Society Series B*, 64, 583–639.
- Zhang, J. E., Shu, J., and Brenner, M., 2010. The new market for volatility trading. *Journal of Futures Markets*, 30(9), 809–833.
- Zhu, S.-P. and Lian, G.-H., 2011. An analytical formula for vix futures and its applications. *Journal of Futures Markets*, 32(2), 166–190.

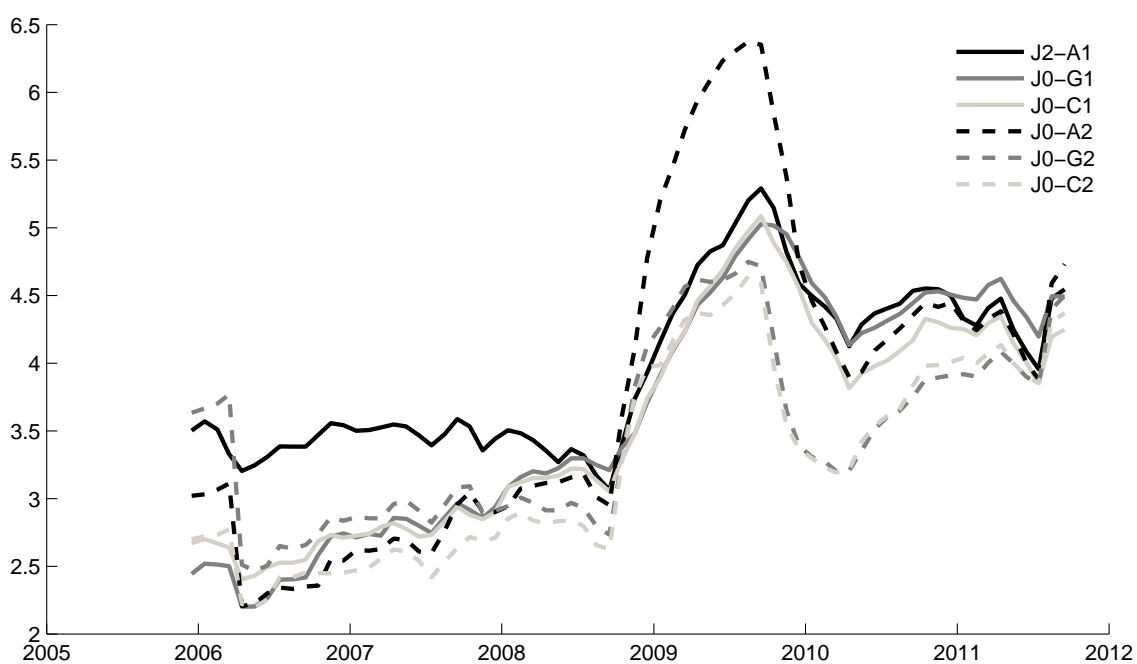


Figure 1: S&amp;P 500 and VIX Term-Structure.



*This figure plots the evolution VIX indices for a maturity of 30 and 360 days, the slope of the volatility term structure as measured by the difference of the 30 and 360 days VIX and daily S&P 500 log returns over the sample period from January 1990 until September 2011.*

Figure 2: **Out-of-sample Pricing Performance.**



*This figure plots evolution of the average monthly RMSE for the out-of-sample pricing exercise for six of the models described in Section 2.. The RMSE is calculated using a rolling window of one year, i.e for each month the RMSE is the average RMSE over the twelve preceding months.*

Table 1: **Descriptive Statistics.**

Maturity (in days)	30	360	30	360	30	360
Variable	Level		Changes in level		Changes in log	
mean	20.28	21.33	0.00	0.00	0.00	0.00
standard deviation	8.25	6.34	1.50	0.53	0.06	0.02
skewness	2.02	1.16	0.19	1.01	0.19	0.58
kurtosis	10.15	4.95	20.98	22.89	9.63	11.17
min	9.17	12.45	-18.58	-4.96	-0.58	-0.14
max	80.76	54.18	14.49	6.67	0.42	0.20

*This table reports descriptive statistics for the 30-day VIX and the 360-day VIX index over the sample period from January 1990 until September 2011.*

Table 2: Estimation Results (One-factor Models).

	J0-A1	J1-A1	J2-A1	J0-G1	J1-G1	J2-G1	J0-C1	J1-C1	J2-C1
$\gamma$	0.5	0.5	0.5	1	1	1	0.764 (0.016)	0.733 (0.015)	0.778 (0.017)
$\kappa_v^P$	4.182 (0.355)	3.714 (0.354)	3.395 (0.328)	4.764 (0.305)	4.330 (0.308)	4.230 (0.3)	4.589 (0.335)	4.098 (0.318)	3.956 (0.335)
$\kappa_v^Q$	4.052 (0.178)	3.861 (0.165)	3.564 (0.191)	3.452 (0.13)	3.244 (0.135)	3.259 (0.139)	3.859 (0.152)	3.722 (0.153)	3.606 (0.162)
$\theta_v^P$	0.042 (0.003)	0.041 (0.004)	0.024 (0.004)	0.029 (0.002)	0.027 (0.003)	0.022 (0.003)	0.035 (0.002)	0.035 (0.003)	0.027 (0.003)
$\sigma_v$	0.370 (0.007)	0.374 (0.008)	0.324 (0.009)	1.791 (0.036)	1.833 (0.045)	1.739 (0.045)	0.873 (0.047)	0.809 (0.042)	0.870 (0.053)
$\rho$	-0.754 (0.009)	-0.790 (0.01)	-0.796 (0.01)	-0.781 (0.009)	-0.819 (0.01)	-0.824 (0.009)	-0.775 (0.009)	-0.812 (0.01)	-0.817 (0.009)
$\lambda$		5.670 (2.056)	5.894 (1.273)		5.107 (1.904)	6.096 (1.382)		5.571 (1.999)	5.179 (1.2)
$\mu_s^P$		0.004 (0.003)	-0.006 (0.003)		0.004 (0.003)	-0.003 (0.003)		0.004 (0.003)	-0.003 (0.003)
$\sigma_s^P$		0.017 (0.003)	0.017 (0.002)		0.019 (0.003)	0.016 (0.002)		0.017 (0.003)	0.017 (0.002)
$\mu_v^P$			0.010 (0.002)			0.005 (0.001)			0.007 (0.001)
DIC	-94617	-95164	-95446	-95231	-95643	-96156	-95149	-95607	-95957

This table reports the parameter estimates for one-factor SV diffusion and jump-diffusion models from daily time-series data of the S&P 500 and 30-day and 360-day VIX indices, over a period from 1990 until 2004. Parameters are estimated by MCMC methods and these estimates are followed by the DIC goodness-of-fit statistic for each model. Point estimates are based on the mean of the posterior distributions, standard deviations of the posterior are given in parenthesis. The models are abbreviated as in Section 2. Parameters are reported as annual decimals.

**Table 3: Estimation Results (Two-factor models).**

	J0-A1	J1-A1	J2-A1	J0-G1	J1-G1	J2-G1	J0-C1	J1-C1	J2-C1
$\gamma_v$	0.5	0.5	0.5	1	1	1	0.960 (0.019)	0.923 (0.019)	0.864 (0.04)
$\gamma_m$	0.5	0.5	0.5	1	1	1	1.178 (0.04)	1.136 (0.042)	1.041 (0.053)
$\kappa_v^P$	4.697 (0.351)	4.307 (0.35)	10.824 (2.79)	5.544 (0.365)	4.846 (0.383)	8.805 (2.663)	5.806 (0.388)	5.125 (0.396)	9.427 (2.763)
$\kappa_v^Q$	4.400 (0.151)	4.026 (0.157)	10.047 (2.568)	4.979 (0.152)	4.336 (0.165)	7.428 (1.822)	5.215 (0.18)	4.630 (0.178)	8.306 (2.098)
$\sigma_v$	0.386 (0.007)	0.382 (0.009)	0.447 (0.038)	2.008 (0.038)	2.024 (0.05)	2.361 (0.23)	1.786 (0.115)	1.617 (0.103)	1.547 (0.157)
$\rho$	-0.774 (0.009)	-0.809 (0.009)	-0.745 (0.031)	-0.797 (0.008)	-0.837 (0.009)	-0.814 (0.016)	-0.795 (0.008)	-0.836 (0.009)	-0.796 (0.021)
$\kappa_m^P$	0.420 (0.257)	0.428 (0.261)	0.224 (0.156)	0.350 (0.233)	0.354 (0.234)	0.170 (0.137)	0.317 (0.218)	0.306 (0.215)	0.158 (0.132)
$\theta_m^P$	0.047 (0.019)	0.046 (0.019)	0.037 (0.024)	0.051 (0.021)	0.049 (0.021)	0.043 (0.027)	0.051 (0.022)	0.050 (0.023)	0.042 (0.027)
$\sigma_m$	0.096 (0.002)	0.098 (0.002)	0.093 (0.004)	0.443 (0.009)	0.460 (0.012)	0.471 (0.022)	0.767 (0.095)	0.699 (0.092)	0.560 (0.096)
$\lambda$		3.883 (1.355)	10.173 (1.997)		5.125 (1.698)	11.411 (2.563)		5.658 (1.835)	11.218 (2.459)
$\mu_s^P$		0.008 (0.004)	-0.005 (0.002)		0.006 (0.003)	-0.003 (0.002)		0.006 (0.003)	-0.003 (0.002)
$\sigma_s^P$		0.019 (0.003)	0.014 (0.001)		0.018 (0.003)	0.014 (0.001)		0.017 (0.002)	0.014 (0.001)
$\mu_v$			0.013 (0.002)			0.005 (0.001)			0.007 (0.001)
DIC	-137122	-136861	-128753	-139128	-138953	-132825	-138729	-138547	-132575

*This table reports the parameter estimates for two-factor volatility models from daily time-series data of the S&P 500 and 30-day and 360-day VIX indices over a period from 1990 until 2004. Parameters are estimated by MCMC methods and these estimates are followed by the DIC goodness-of-fit statistic for each model. Point estimates are based on the mean of the posterior distributions, standard deviations of the posterior are given in parenthesis. The models are abbreviated as in Section 2. Parameters are reported as annual decimals.*

Table 4: In-Sample Fit to VIX Term Structure.

	J0-A1	J1-A1	J2-A1	J0-G1	J1-G1	J2-G1	J0-C1	J1-C1	J2-C1
$\sigma_\varepsilon(30)$	0.024 (0.001)	0.025 (0.001)	0.028 (0.002)	0.027 (0.001)	0.029 (0.001)	0.028 (0.001)	0.025 (0.001)	0.026 (0.001)	0.027 (0.001)
$\rho_\varepsilon(30)$	0.019 (0.010)	0.025 (0.011)	0.054 (0.028)	0.028 (0.01)	0.046 (0.016)	0.054 (0.021)	0.018 (0.01)	0.028 (0.013)	0.045 (0.023)
$\sigma_\varepsilon(360)$	0.013 ( $<0.001$ )	0.013 ( $<0.001$ )	0.013 ( $<0.001$ )	0.013 ( $<0.001$ )	0.013 ( $<0.001$ )	0.013 ( $<0.001$ )	0.013 ( $<0.001$ )	0.013 ( $<0.001$ )	0.013 ( $<0.001$ )
$\rho_\varepsilon(360)$	0.982 (0.001)	0.980 (0.002)	0.981 (0.002)	0.978 (0.002)	0.977 (0.002)	0.978 (0.002)	0.981 (0.001)	0.979 (0.002)	0.980 (0.002)
	J0-A2	J1-A2	J2-A2	J0-G2	J1-G2	J2-G2	J0-C2	J1-C2	J2-C2
$\sigma_\varepsilon(30)$	0.024 (0.001)	0.027 (0.002)	0.020 (0.002)	0.026 (0.001)	0.028 (0.001)	0.022 (0.002)	0.025 (0.001)	0.027 (0.001)	0.020 (0.002)
$\rho_\varepsilon(30)$	0.022 (0.010)	0.050 (0.024)	0.032 (0.025)	0.025 (0.010)	0.047 (0.015)	0.037 (0.020)	0.023 (0.010)	0.040 (0.014)	0.028 (0.019)
$\sigma_\varepsilon(360)$	0.006 ( $<0.001$ )	0.006 ( $<0.001$ )	0.015 (0.002)	0.005 ( $<0.001$ )	0.005 ( $<0.001$ )	0.013 (0.002)	0.005 ( $<0.001$ )	0.006 (0.001)	0.014 (0.002)
$\rho_\varepsilon(360)$	0.009 (0.010)	0.012 (0.012)	0.302 (0.182)	0.002 (0.010)	0.010 (0.014)	0.217 (0.139)	0.003 (0.010)	0.014 (0.016)	0.234 (0.147)

This table reports the in-sample fit of the error dynamics (12) for one-factor SV models (upper part) and for two-factor SV models (lower part) from daily time-series data on the S&P 500 and 30-day and 360-day VIX indices over a period from 1990 until 2004. Parameters are estimated by MCMC methods. Point estimates are based on the mean of the posterior distributions, standard deviations of the posterior are given in parenthesis. The models are abbreviated as in Section 2.

Table 5: **Out-of-Sample Option Pricing Errors.**

One-factor Models	J0-A1	J1-A1	J2-A1	J0-G1	J1-G1	J2-G1	J0-C1	J1-C1	J2-C1
RMSE (%)	4.709	4.329	4.271	3.925	4.419	3.967	3.845	4.258	3.950
Two-factor Models	J0-A2	J1-A2	J2-A2	J0-G2	J1-G2	J2-G2	J0-C2	J1-C2	J2-C2
RMSE (%)	4.512	4.493	4.572	4.118	4.089	4.098	3.772	3.905	3.947

*This table reports the bias and efficiency of out-of-sample option pricing errors for the entire out-of-sample period from January 2005 until September 2011. The models are abbreviated as in Section 2. The statistics are based on 143,019 Wednesday close prices for options with moneyness between 0.8 and 1.1 and time to maturity between 7 and 550 days.*

Table 6: **Out-of-Sample Option Pricing Errors (Maturity).**

One-factor Models	J0-A1	J1-A1	J2-A1	J0-G1	J1-G1	J2-G1	J0-C1	J1-C1	J2-C1
<b>Maturity</b>									
7 to 90 (# 91,260)	5.249	4.734	4.685	3.866	4.663	4.096	3.944	4.616	4.189
91 to 180 (# 21,453)	2.881	2.843	2.813	2.986	2.917	2.802	2.829	2.736	2.728
181 to 365 (# 21,522)	3.823	3.748	3.629	4.379	4.294	4.045	3.960	3.824	3.753
366 to 550 (# 8,784)	4.324	4.260	4.167	5.170	5.103	4.765	4.596	4.437	4.354
Two-factor Models	J0-A2	J1-A2	J2-A2	J0-G2	J1-G2	J2-G2	J0-C2	J1-C2	J2-C2
<b>Maturity</b>									
7 to 90 (# 91,260)	5.118	5.115	5.031	4.601	4.593	4.563	4.047	4.250	4.204
91 to 180 (# 21,453)	2.906	2.888	3.474	2.642	2.562	2.779	2.777	2.735	3.121
181 to 365 (# 21,522)	3.279	3.187	3.730	3.261	3.169	3.276	3.418	3.398	3.616
366 to 550 (# 8,784)	3.505	3.392	3.712	3.607	3.530	3.467	3.757	3.748	3.771

*This table reports out-of-sample RMSE pricing errors for the out-of-sample period from January 2005 until September 2011, with errors stratified by the call option maturity. The models are abbreviated as in Section 2.*



Table 7: **Out-of-Sample Option Pricing Errors (Moneyness).**

One-factor Models	J0-A1	J1-A1	J2-A1	J0-G1	J1-G1	J2-G1	J0-C1	J1-C1	J2-C1
<b>Moneyness</b>									
0.8 to 0.9 (# 40,189)	6.135	4.816	4.505	4.486	5.832	4.590	4.290	5.251	4.401
0.9 to 0.97 (# 37,294)	2.762	2.686	2.819	2.947	2.893	2.758	2.693	2.707	2.638
0.97 to 1.03 (# 35,791)	3.047	3.112	3.054	2.846	2.850	2.820	2.839	2.885	2.886
1.03 to 1.1 (# 29,749)	5.918	6.171	6.254	5.122	5.259	5.301	5.235	5.540	5.485
Two-factor Models	J0-A2	J1-A2	J2-A2	J0-G2	J1-G2	J2-G2	J0-C2	J1-C2	J2-C2
<b>Moneyness</b>									
0.8 to 0.9 (# 40,189)	4.529	4.523	4.019	4.710	4.616	4.699	3.880	4.180	3.847
0.9 to 0.97 (# 37,294)	3.110	3.084	3.195	2.762	2.622	2.488	2.490	2.521	2.621
0.97 to 1.03 (# 35,791)	3.424	3.467	3.641	3.031	3.019	3.031	2.931	2.952	3.126
1.03 to 1.1 (# 29,749)	6.627	6.558	7.064	5.562	5.658	5.666	5.474	5.591	5.876

*This table reports out-of-sample RMSE pricing errors for the out-of-sample period from January 2005 until September 2011, with errors stratified by the call option moneyness. The models are abbreviated as in Section 2.*