Continuous-time VIX dynamics: On the role of stochastic volatility of volatility

Andreas Kaeck, Carol Alexander

Abstract

This paper examines the ability of several different continuous-time one- and two-factor jump-diffusion models to capture the dynamics of the VIX volatility index for the period between 1990 and 2010. For the one-factor models we study affine and non-affine specifications, possibly augmented with jumps. Jumps in one-factor models occur frequently, but add surprisingly little to the ability of the models to explain the dynamics of the VIX. We present a stochastic volatility of volatility model that can explain all the time-series characteristics of the VIX studied in this paper. Extensions demonstrate that sudden jumps in the VIX are more likely during tranquil periods and the days when jumps occur coincide with major political or economic events. Using several statistical and operational metrics we find that non-affine one-factor models outperform their affine counterparts and modeling the log of the index is superior to modeling the VIX level directly.

1. Introduction

As a measure of volatility implied in traded equity index option prices, volatility indices have attracted research for almost a decade. The diverse problems being investigated include: the construction methodology; their use in constructing trading strategies; their role in constructing an information content regarding future volatility; and their use in constructing volatility indices. The liquidity of these contracts has increased dramatically since the international banking crisis of 2008 and a wide range of futures, options and swaps is now available for trading. Market participants use these instruments for diversification, hedging options and pure speculation. To this end, several pricing models have been considered (e.g. Grubbichler & Longstaff, 1996; Whaley, 1993). Empirical evidence regarding the data generating process of volatility indices is, however, still scarce. To date, the only comparative study of alternative data generating processes is Dotsis et al. (2007) who investigate the performance of several affine one-factor models using a sample from 1997 to 2004. They find that a Merton-type jump process outperforms other models for a wide range of different volatility indices. Extensions of some of the models are also considered in Dotsis et al. (2010). In general, there is little disagreement in the literature regarding some important characteristics of volatility, such as the need for a mean-reverting process to account for a long-term equilibrium value. There is also evidence that volatility jumps constitute a relatively large fraction of the variability of volatility indices. Psychoyios et al. (2010) argue that these jumps are an important feature and show that omitting them from the data generating process can lead to considerable differences in VIX option prices and hedge ratios.

Jumps in volatility may also be important for modeling equity index returns, as for instance in Eraker, Johannes, and Polson (2003). Yet

2 Dotsis et al. (2007) point out that this feature is only of second order importance as the best performing model in their study is a Merton-type jump process without a mean-reversion component. Modeling volatility with this process over a long-time horizon is however problematic as in this model volatility tends to either zero or infinity in the long run.

Keywords:
VIX
Volatility indices
Jumps
Stochastic volatility of volatility

© 2013 Published by Elsevier Inc.
curiously, there is a large discrepancy between the volatility jump intensities estimated using two-factor models on equity index time series and one-factor models on volatility index time series. The most important example is the difference between the S&P 500 index and its volatility index VIX. Eraker et al. (2003) estimate about 1.5 volatility jumps per year when based on equity index data, yet Dotis et al. (2007) estimate between 28 to 100 volatility jumps (depending on the model) using the VIX. Although the estimates are not directly comparable due to the different sample periods and the different modeling approaches, their huge differences are still puzzling.

This paper makes several novel contributions to understanding the continuous-time dynamics of volatility indices. In particular, we extend the data and the methodology of other empirical research on volatility index dynamics in four important ways. Firstly, we study the VIX over a long-time horizon of more than 20 years which includes the recent banking and credit crisis. Using a long time series covering several periods of market distress is essential if we are to uncover all dimensions of its historical behavior. Moreover, we have observed several different market regimes over the last two decades, and we shall seek a model that can explain the VIX dynamics during all types of market circumstances. The recent crisis period is of particular importance, as this prolonged period of high volatility revealed vital information regarding the extreme behavior of volatility. Understanding this behavior is particularly important, as it influences numerous aspects of risk and portfolio management.

Secondly, we depart from standard affine model specifications and study the dependence of the diffusion part on the level of the index. Non-affine models have recently attracted much attention, for example Christoffersen, Jacobs, and Mimouni (2010) find that non-affine specifications outperform affine processes in an equity index option pricing framework and Chourdakis and Dotis (2011) confirm these findings using a joint-time series of the VIX and the underlying S&P 500 index returns. In our context, the chief motivation to study these models is that a stronger dependence of the diffusion term on the VIX level might decrease the jump intensity of the models. Extremely high jump intensities may be problematic because one loses the economic reasoning that jumps cover large, unexpected movements in the time-series. The estimation of non-affine models is, however, more difficult to handle, as discrete-time transition probabilities or characteristic functions are generally unavailable in closed form. Our approach includes the estimation of these processes with a Markov-chain-Monte-Carlo sampler using a data augmentation technique as in Jones (1998). This procedure allows us to study a wide range of processes, affine and otherwise, within the same econometric framework.

The third and perhaps the most important contribution is the extension of existing volatility dynamic models to the case of stochastic volatility (stochastic vol-of-vol hereafter). This feature has, to our knowledge, not been studied for the time-series of volatility indices before, but it yields very attractive properties: increasing variability can be modeled as a persistent vol-of-vol component rather than indirectly via the VIX. Mean reverting processes are now an accepted starting point for the deviation from this value the stronger the drift of the process pulls the process back toward its long-term mean. Constant and zero drift components have been criticized for ignoring this feature and hence are – at least in the long run – regarded an unrealistic description of volatility. Mean reverting processes are now an accepted starting point for volatility and variance modeling.

The diffusion term of a continuous-time process is often chosen so that the model falls into the class of affine processes. To model the VIX and other volatility indices, Dotis et al. (2007) rely on the square-root and a Merton-type jump model for volatility and Psychoyios et al. (2010) also consider an Ornstein–Uhlenbeck process to model the log of VIX. In this paper, we study several extensions of these models. For modeling both VIX and its log process, we allow the diffusion function to be proportional to the process. Variants of these models have been successfully applied in other contexts, such as option pricing or spot index modeling (see Chernov, Gallant, Ghysels, & Tauchen, 2003 or Christoffersen et al., 2010). Especially for option pricing applications researchers often favor square-root specifications, as they retain the concept of predictive p-values to study a wide range of characteristics of all the processes under consideration. This is crucial, as previous studies focused mainly on the relative performance of the models. We find that the stochastic vol-of-vol model generates dynamics that are, of all the models considered, most closely in line with the observed VIX time series. Finally we provide empirical evidence using a scenario analysis exercise.

The new CBOE volatility index construction methodology provides a close link between instantaneous variance and the VIX index. In pure diffusion models, the squared VIX is the risk-neutral expectation of integrated variance over a 30-day horizon. This fact is very useful not only to establish a theoretical link between the two quantities, but also to model S&P 500 index options and derivatives written on the VIX in a consistent manner (see Sepp, 2008; Zhang & Zhu, 2006 or Zhu & Lian, 2012). In this paper, we model the VIX index directly without explicitly specifying instantaneous variance dynamics. While it would be desirable to model various derivative markets with the same underlying stochastic process, whether standard option pricing models can provide a reasonable fit to several markets at the same time is an open research question. For instance, Broadie, Chernov, and Johannes (2007) report relatively large option pricing errors for the Heston model when time-series consistency is imposed on the structural parameters. Much in the same way it is not evident if equity and VIX derivative markets can be unified with one stable parametric model. In addition, modeling the VIX via the instantaneous variance process would require to specify both the real-world and the risk-neutral dynamics. Our goal is to examine alternative processes for the VIX index directly and to provide evidence regarding the type of process that is needed to explain its empirical characteristics.

We proceed as follows: Section 2 introduces the affine and non-affine one-factor models used; Section 3 describes our econometric estimation methodology. Section 4 provides details on the data set. In Section 5 we provide estimation results for various alternative one-factor processes. Section 6 introduces and presents results for the stochastic vol-of-vol model. We provide a risk management application in Section 7 and Section 8 concludes.

2. Model specifications

Most models proposed for describing volatility or variance dynamics agree on its mean-reverting nature. This feature reflects the belief that, although volatility can temporarily fluctuate widely, it will never wander away too much from its long-term equilibrium value. The stronger the deviation from this value the stronger the drift of the process pulls the process back toward its long-term mean. Constant and zero drift components have been criticized for ignoring this feature and hence are – at least in the long run – regarded an unrealistic description of volatility. Mean reverting processes are now an accepted starting point for volatility and variance modeling.

The diffusion term of a continuous-time process is often chosen so that the model falls into the class of affine processes. To model the VIX and other volatility indices, Dotis et al. (2007) rely on the square-root and a Merton-type jump model for volatility and Psychoyios et al. (2010) also consider an Ornstein–Uhlenbeck process to model the log of VIX. In this paper, we study several extensions of these models. For modeling both VIX and its log process, we allow the diffusion function to be proportional to the process. Variants of these models have been successfully applied in other contexts, such as option pricing or spot index modeling (see Chernov, Gallant, Ghysels, & Tauchen, 2003 or Christoffersen et al., 2010). Especially for option pricing applications researchers often favor square-root specifications, as they retain

---

3 Merino and Sentana (in press) also find that a stochastic vol-of-vol model has favourable VIX option pricing performance, however their model differs from ours as they model the vol-of-vol process with a Levy process that is independent of the VIX index dynamics.

4 Only few exceptions with non-reverting or zero drift components have been proposed in the literature, the SABR model of Hagan, Kumar, Lesniewski, and Woodward (2002) and the Hull and White (1987) model being the most popular.
tractability with analytic pricing formulae for vanilla options, and as such they are relatively easy to calibrate to the market prices of these options.

Another feature that has been found essential in volatility modeling is the inclusion of jumps. Eraker et al. (2003) (using return data) and Brodie et al. (2007) (using both return and option data) find severe misspecifications when jumps in volatility are omitted and document the improvement of variance specifications with exponential upward jumps. Dotsis et al. (2007) report similar results for volatility indices. Whereas previously-mentioned research is based on the assumption that jumps occur as i.i.d. random variables, there is also evidence that jumps in VIX occur more frequently in high volatility regimes (see Psychou et al., 2010).

In order to assess the importance of the characteristics outlined, we employ a general one-factor model in our empirical analysis that accommodates all of the features previously mentioned. Extensions to these models will be considered in Section 6. First we study models that are nested in the following specification:

\[ dX_t = \kappa (\theta - X_t) dt + \sigma X_t^\delta dW_t + Z_t dJ_t, \]

where \( X \) either denotes the value of the volatility index or its logarithm, \( \kappa \) is the speed of mean reversion, \( \theta \) determines the long term value of the process and \( \sigma \) is a constant in the diffusion term. The exponent \( \delta \) is set either to one-half or one for the level of the index, and to zero or one for the log process. Note that if \( b = 1 \) in the log process, VIX is bounded from below by one whereas the lower bound is zero in the other models. As remarked by Chernov et al. (2003), this is a very mild restriction for yearly volatility.5

In terms of jump distributions we assume that \( J \) is a Poisson process with time varying intensity \( \lambda_0 + \lambda_0 X_t \). For the jump sizes we consider two alternatives. Firstly, we employ an exponentially distributed jump size, as this assumption is commonly applied to the variance in equity markets. The exponential distribution has support on the positive real axis, so it allows for upward jumps only, which guarantees that the process does not jump to a negative value. The distribution is parsimonious with only one parameter \( \lambda_0 \), representing both the expectation and the volatility of the jump size, to estimate. We apply this jump size distribution to all models except for the log volatility model with \( b = 0 \), for which we use normally distributed jump sizes with mean \( \mu_J \) and standard deviation \( \sigma_J \) because the support of this model is not restricted to positive numbers and the log volatility may become negative.6

3. Econometric methodology

3.1. Estimation of jump-diffusion models

Several estimation techniques for jump-diffusion processes have been proposed in the literature. In the context of volatility indices, Dotsis et al. (2007) use conditional maximum likelihood methods to estimate the structural parameters of several alternative processes for six different volatility indices. Psychou et al. (2010) adopt the same methodology to the VIX and also include state-dependent jump diffusion models. In this paper, we adopt a Bayesian Markov-chain-Monte-Carlo (MCMC) algorithm because this estimation technique has several advantages over other approaches, particularly for the models we consider.7

Firstly, it provides estimates not only for structural parameters, but also for unobservable latent variables such as the jump times and jump sizes. These latent parameter estimates provide valuable information for testing the model and shed light on whether key assumptions of the model are reflected in our estimates. Secondly, our algorithm allows one to handle non-affine models for which closed-form transition densities or characteristic functions are unavailable.

The center of interest for our analysis is the joint distribution of parameters and latent variables conditional on the observed data. In Bayesian statistics, this distribution is termed the posterior density and is given by

\[ p(\theta, Z, J | X) \propto p(X | \theta, Z, J) p(\theta) p(Z, J) \]

where the first density on the right is the likelihood of the observed data conditional on model parameters and latent state variables and the second density denotes the prior beliefs about parameters and latent state variables, not conditional on the data. The vector \( \theta \) collects all structural parameters, and \( Z \) and \( J \) collect all jump sizes, jump times and VIX (or \( \log(\text{VIX}) \)) observations respectively.

Knowing the posterior density we can estimate point 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 uninformative, hence our parameter estimates are driven by the information in the data and not the prior.8 But there remain two questions to address: how to determine the likelihood, because a closed-form density can only be obtained for some models of the affine class, and how to recover the posterior density.

To obtain a closed-form likelihood we can approximate the evolution of the continuous-time process for the volatility index by a first-order Euler discretization. Therefore between two time steps the process evolves according to

\[ X_{t+1} = X_t + \sum_{i=1}^{b} \left[ \kappa (\theta - X_t) h + \sigma X_t^\delta \epsilon_{t+1,i} + Z_{t+1,i} \right] \]

where \( h \) denotes the discretization step, \( \epsilon_i \) denote standard normal variates and the jump process is discretized by assuming that the event \( J_{t+1,i} = 1 \) occurs with probability \( \epsilon_{t+1,i} \). This approximation converges (under some regularity conditions) to the true continuous-time process as \( h \rightarrow 0 \). Therefore choosing \( h \) to be small should lead to a negligible discretization bias. But in reality the frequency of the observed data cannot be chosen by the researcher. In our case data are recorded daily and so the discretization bias could be substantial, depending on the structural parameters of the model.9

A great advantage of the MCMC approach is that it allows one to augment the observed data with unobserved, high-frequency observations, a technique that has been applied to continuous-time diffusion and jump-diffusion models in Jones (1998) and Eraker (2001). This way, we treat data points between two observations as unobserved or missing data. Hence, even if the data set only includes daily values for the VIX, we can estimate the parameters of the continuous-time process accurately by choosing \( h \) small and augmenting the observed data. Here there are two practical issues that need addressing. Firstly, decreasing \( h \) leads to increasing computational cost and it also increases the parameters to be estimated substantially. And secondly, the inclusion of many data points makes it more difficult for the algorithm to filter out jump times and jump sizes because the signaling effect of a large daily observation becomes weaker. Throughout this paper we use \( h = 0.25 \).

---

5 To avoid this one could model not the VIX directly, but its value minus this lower bound.

6 In fact, we have also estimated all models with both normally and exponentially distributed jump size, so that we may get the effect of this assumption on the model performance. Since in some models the normal distribution can lead to negative VIX values and we found only little improvements from this more general jump size distribution, we report only results for one distribution in each model. All of our qualitative conclusions are robust with respect to changing this jump size distribution.

7 MCMC methods in financial econometrics were pioneered in Jacquier, Polson, and Rossi (1994).

8 Details about these distributions are provided upon request.

9 The discretization of the jump part, especially, may lead to a large bias because daily observations allow no more than one jump per day. According to the results in Dotsis et al. (2007) volatility indices can jump far too frequently for this to be negligible. However, if the jump intensity is much lower, as in Eraker et al. (2003), a daily discretization does not introduce any discernible error.
Note that although we generate a distribution of each augmented data point, we have no interest in the density of $X^n$ itself; it is used only to decrease the discretization bias.

The second question, of recovering the posterior density, is dealt with by applying a Gibbs sampler (Geman & Geman, 1984). This approach achieves the goal of simulating from the 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 (under mild regularity conditions) is the posterior density. More precisely, step g in the Markov chain consists of:

1. Draw the latent variables: $p\{X^{(g)}|\Theta^{(g-1)}, Z^{(g-1)}\}
   \quad p\{Z^{(g)}|\Theta^{(g-1)}, X^{(g-1)}\}
   \quad p\{\Theta^{(g)}|X^{(g-1)}, Z^{(g-1)}\}$

2. Draw structural parameters: $p\{\Theta^{(g)}|X^{(g-1)}, Z^{(g)}, J\}$.

The latent state vectors and structural parameters can be further divided into blocks, so that we only need to draw from one-dimensional distributions. Some of the univariate distributions are of unknown form and we use a Metropolis algorithm for these.\(^{10}\)

3.2. Model specification tests

In order to test different specifications we employ a simple but powerful test procedure. Taking a random draw of the vector of structural parameters from the posterior distribution, we use this to simulate a trajectory of the same sample size as the original VIX time series. Given this trajectory, we calculate several sample statistics and compare them with the observed sample statistics obtained from the original VIX time series. Applying this procedure several thousand times we obtain a distribution for each statistic and for each model under consideration. Finally, for each statistic and each model, we compute the probability associated with the value of the statistic given by the observed VIX time series under the model’s distribution for the statistic. This p-value reveals how likely the observed value of the statistic is, according to the model. Very high or low p-values convey the model’s inability to generate the observed data. For more details on this type of model specification testing procedure we refer to Rubin (1984), Meng (1994), Gelman, Meng, and Stern (1996) and Bayarri and Berger (2000).

It is common to use higher order moments to discriminate between alternative specifications. For example, if the estimated models are realistic descriptions of VIX dynamics, then in repeated simulations the models should create kurtosis levels similar to the observed. We shall choose a wide range of statistics that we deem important for modeling volatility indices, including:\(^ {11}\)

- The descriptive statistics in Table 1 below except for the unconditional mean (because with a mean-reverting process the mean only indicates whether the start value is below or above the last simulated value and this is of no interest). That is we opt for standard deviation (stddev), skewness (skew) and kurtosis (kurt) and the minimum (min) and maximum (max) of the process. Note that these statistics indicate whether a model can capture the standardized moments up to order four, as well as the extreme movements of the VIX.
- Statistics on the highest positive and negative changes in the index (minjump and maxjump), the average over the 10 largest positive changes (avgmaj10) and the average over the 10 largest negative changes (titavmin10). These statistics shed light on whether the model can replicate the observed outliers.\(^ {12}\)
- In order to investigate the clustering of the outliers we use the month (20 trading days) with the highest sum of absolute changes in the process (absmax20). Likewise we report the statistic for the month with the least absolute changes (absmin20). Taken together these two statistics reflect our belief that the model should be able to reproduce periods of low activity and periods of high uncertainty in the level of the VIX.
- Finally, we report various percentiles of the estimated unconditional distribution of daily changes in the VIX. The percentiles are denoted by percNUM where NUM indicates the percentage level, and they indicate whether the model can replicate the observed unconditional density.

To simulate the continuous-time processes we use the same time-discretization as we have employed for the estimation of the processes. Furthermore, we start each simulation at the long-term mean value of the VIX and use 50,000 trajectories to calculate the p-values.

This test procedure has several advantages over simple in-sample fit statistics (most of which do not, in any case, apply to the Bayesian framework we use). Firstly, it allows us to detect exactly which characteristics of the VIX a model struggles to reproduce. Secondly, it allows us to compare the models in both a relative and an absolute sense. That is, as well as comparing the performance of competing models, our procedure also indicates whether each model provides a good or bad description of the observed VIX dynamics. Thirdly, it takes parameter uncertainty into account because it draws the structural parameters randomly from the posterior density.

### Table 1

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>Std dev</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.32</td>
<td>9.31</td>
<td>80.86</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First difference</td>
<td>0.003</td>
<td>1.512</td>
<td>0.427</td>
<td>21.819</td>
<td>-17.36</td>
<td>16.54</td>
</tr>
</tbody>
</table>

\(^{10}\) Details about this algorithm are provided upon request. A standard reference including a wide range of Metropolis algorithms is Robert and Casella (2004).

\(^{11}\) Very similar statistics have also been used for testing equity index dynamics by Kaeck and Alexander (in press).

\(^{12}\) The label of the first two statistics includes the term ‘jump’ but this does not imply that they test for the presence of a jump. In fact, all four statistics in this group only capture the ability of a model to explain outliers, without attributing any cause. They could be the result of jumps, or of particularly extreme returns generated by a pure diffusion model.
market crises. The first such period in our sample corresponds to the outbreak of the first Gulf War in August 1990, when the VIX exceeded 30% for several months. Following this, markets stayed calm for a couple of years until July 1997. During this tranquil volatility regime the VIX only temporarily exceeded 20%. With the Asian crisis in 1997 we entered a sustained period of high uncertainty in equity markets. Several financial and political events contributed to this: the Long Term Capital Management bailout in 1998, the bursting of the Dot-Com bubble in 2000 and the 9/11 terror attacks leading to the second Gulf War in 2001. In 2003 VIX levels begin a long downward trend as equity markets entered another tranquil period which prevailed until 2007. Then, after the first signs of a looming economic crisis surfaced, VIX rose again. Following the Lehman Brothers collapse in September 2008 it appeared to jump up, to an all-time high of over 80%. Before this such high levels of implied volatility had only been observed during the global equity market crash of 1987, which was before the VIX existed. Equity markets returned to around 20% volatility in 2009, but then with the Greek crisis in May 2010, at the end of the sample, the VIX again appeared to jump up, to around 40%.

Table 1 reports descriptive statistics for the VIX. From a modeling perspective the most interesting and challenging characteristic are some huge jumps in the index, indicated by the very large min and max values of the first difference. Movements of about 15% per day (about 10 standard deviations!) will pose a challenge to any model trying to describe the evolution of the indices. Interestingly downward jumps can be of an even higher magnitude and we will discuss this issue further below.

5. Estimation results

5.1. Jump-diffusion models on the VIX level

First we focus on the jump-diffusion models for the VIX level with $b = 0.5$, which are reported in the left section of Table 2. Starting with the pure diffusion model in the first column, we estimate a speed of mean reversion $\kappa$ of 0.016 which corresponds to a characteristic time to mean revert of $1/0.016 = 63$ days. One minus this parameter is approximately the first-order autocorrelation of the time series, hence our results imply that volatility is highly persistent. The long-term volatility value $\theta$ is about 20.5% which is close to the unconditional mean of the process in Table 1. Our parameter estimate for $\alpha$ is 0.289. 13

Several interesting features arise when considering the exponential jump models in columns 2, where $\lambda_1 = 0$ so that jump intensities are independent of the level of the VIX, and column 3 where $\lambda_0 = 0$ but jump intensities depend on the level of the VIX. 14 Firstly, the inclusion of jumps increases the speed of mean reversion considerably, to 0.037 when $\lambda_1 = 0$ and 0.051 when $\lambda_0 = 0$. A possible explanation is that the drift of the process tries to compensate for omitted downward jumps, so that when volatility is exceptionally high the process can create larger downward moves with an increased $\kappa$ estimate. Furthermore, in the jump models the estimates for the second drift parameter $\theta$ drop to about 12–14%, a result that is expected because $\theta$ carries a different interpretation once jumps are included. To obtain the long-term volatility we have to adjust $\theta$ by the effect of jumps and our estimation results imply long-term volatility levels of approximately 21%, similar to the pure diffusion model. As expected the parameter $\alpha$ decreases in all jump models since part of the variation in the VIX is now explained by the jump component.

When jump probabilities are assumed to be independent of the VIX level, a jump occurs with a likelihood of 0.107 per day. A parameter of this magnitude implies about 27 jumps per year, hence such events may be far more frequent than for many other financial variables such as stock prices or interest rates. An average-sized jump is 2.38 VIX points. Jump occurrence in the models with state-dependent jumps is higher, with average jump probabilities of about 26%. 15 As we estimate more jumps in this case, the average jump size decreases to only 1.58 VIX points.

13 Note that this model was previously studied in Dotis et al. (2007) but these authors used VIX data from the generally volatile period from October 1997 to March 2004 so our results are not directly comparable. Not surprisingly, the parameter estimates in Dotis et al. (2007) imply more rapidly moving processes than ours: they estimate a (yearly) speed of mean reversion of 9.02 (whereas our yearly equivalent is 4.03) and a long-term volatility level of 24.5%.

14 We have also estimated all models with $\lambda_0$ and $\lambda_1$ being simultaneously different from zero but these results are omitted for expositional clarity. The parameter estimates for these models reveal that jump probabilities are mainly driven by the state-dependent jump part as $\lambda_0$ is close to zero. Therefore, the evidence appears to point toward state-dependent jumps. We return to this observation later on.

15 This estimate is based on an average VIX level of about 20%. 

---

5. Estimation results

5.1. Jump-diffusion models on the VIX level

First we focus on the jump-diffusion models for the VIX level with $b = 0.5$, which are reported in the left section of Table 2. Starting with the pure diffusion model in the first column, we estimate a speed of mean reversion $\kappa$ of 0.016 which corresponds to a characteristic time to mean revert of $1/0.016 = 63$ days. One minus this parameter is approximately the first-order autocorrelation of the time series, hence our results imply that volatility is highly persistent. The long-term volatility value $\theta$ is about 20.5% which is close to the unconditional mean of the process in Table 1. Our parameter estimate for $\alpha$ is 0.289. 13

Several interesting features arise when considering the exponential jump models in columns 2, where $\lambda_1 = 0$ so that jump intensities are independent of the level of the VIX, and column 3 where $\lambda_0 = 0$ but jump intensities depend on the level of the VIX. 14 Firstly, the inclusion of jumps increases the speed of mean reversion considerably, to 0.037 when $\lambda_1 = 0$ and 0.051 when $\lambda_0 = 0$. A possible explanation is that the drift of the process tries to compensate for omitted downward jumps, so that when volatility is exceptionally high the process can create larger downward moves with an increased $\kappa$ estimate. Furthermore, in the jump models the estimates for the second drift parameter $\theta$ drop to about 12–14%, a result that is expected because $\theta$ carries a different interpretation once jumps are included. To obtain the long-term volatility we have to adjust $\theta$ by the effect of jumps and our estimation results imply long-term volatility levels of approximately 21%, similar to the pure diffusion model. As expected the parameter $\alpha$ decreases in all jump models since part of the variation in the VIX is now explained by the jump component.

When jump probabilities are assumed to be independent of the VIX level, a jump occurs with a likelihood of 0.107 per day. A parameter of this magnitude implies about 27 jumps per year, hence such events may be far more frequent than for many other financial variables such as stock prices or interest rates. An average-sized jump is 2.38 VIX points. Jump occurrence in the models with state-dependent jumps is higher, with average jump probabilities of about 26%. 15 As we estimate more jumps in this case, the average jump size decreases to only 1.58 VIX points.

13 Note that this model was previously studied in Dotis et al. (2007) but these authors used VIX data from the generally volatile period from October 1997 to March 2004 so our results are not directly comparable. Not surprisingly, the parameter estimates in Dotis et al. (2007) imply more rapidly moving processes than ours: they estimate a (yearly) speed of mean reversion of 9.02 (whereas our yearly equivalent is 4.03) and a long-term volatility level of 24.5%.

14 We have also estimated all models with $\lambda_0$ and $\lambda_1$ being simultaneously different from zero but these results are omitted for expositional clarity. The parameter estimates for these models reveal that jump probabilities are mainly driven by the state-dependent jump part as $\lambda_0$ is close to zero. Therefore, the evidence appears to point toward state-dependent jumps. We return to this observation later on.

15 This estimate is based on an average VIX level of about 20%.
We now turn to the non-affine models with $b = 1$ in the right half of Table 2. There are several interesting results. Firstly the speed of mean reversion $\kappa$ is smaller than in the square-root models. This possibly stems from the fact that the diffusion term, through its stronger dependence on the level of the VIX during high-volatility regimes, can create larger downward jumps and this requires a less rapidly mean-reverting process. The long-term level of the VIX is, as in the square-root model, consistent with its unconditional mean. The diffusion parameter $\sigma$, however, is not comparable with previously studied models and its estimates range from 0.048 to 0.062. State-independent exponentially distributed jumps occur with a likelihood of 0.082 per day and state-dependent jumps are again more likely than state-independent jumps, but they occur only about half as often as in the square-root model class. This has an effect on estimated jump-sizes, where we find that jumps in the non-affine models are more rare events, but their impact is greater and all jump size estimates are larger than in the square-root models. Overall, the jump intensities in non-affine models are still relatively high.

Table 3 provides results from our simulation experiments. These show that the square-root diffusion model is fundamentally incapable of producing realistic data as it fails to generate statistics similar to the observed values for almost every statistic we use. Some of the results are improved when jumps are added, for example using state-independent jumps the standard deviation and the kurtosis of the data yield more realistic values. Nevertheless, overall the square-root model with or without jumps does a very poor job of explaining the characteristics of the VIX. The results for the non-affine specification are more encouraging. Whereas several statistics could not even be produced once in our 50,000 simulations for the square-root diffusion, the non-affine specification does a far better job of matching the observed characteristics of the VIX. However, in absolute terms the non-affine models, with or without jumps, are still severely misspecified. Again, there appears to be little benefit from introducing jumps into the models as the models especially fail to reproduce the statistics that are linked to the jump behavior of the VIX.

### Table 2
Parameter estimates (level models). This table reports the estimates for the structural parameters. The posterior mean is reported as the point estimate, posterior standard deviations and 5%–95% posterior intervals are reported in brackets.

<table>
<thead>
<tr>
<th>Jump distribution</th>
<th>VIX with $b = 0.5$</th>
<th>VIX with $b = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Exp</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.016</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>20.496</td>
<td>13.747</td>
</tr>
<tr>
<td></td>
<td>(1.187)</td>
<td>(0.567)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>[18.667, 22.544]</td>
<td>[12.806, 14.665]</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.289</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>0.107</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2.736</td>
<td>1.584</td>
</tr>
<tr>
<td></td>
<td>(0.185)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>2.104</td>
<td>2.104</td>
</tr>
<tr>
<td></td>
<td>(2.706)</td>
<td>(1.759)</td>
</tr>
</tbody>
</table>

---

### Table 3
Simulation results (level models). This table reports the p-values for all the statistics described in Section 3. The closer these values are to 1 or 0 the greater the degree of model misspecification.

<table>
<thead>
<tr>
<th>Jump distribution</th>
<th>Data</th>
<th>VIX with $b = 0.5$</th>
<th>VIX with $b = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Exp</td>
<td>No</td>
</tr>
<tr>
<td>$\lambda_0$</td>
<td>0.9988</td>
<td>0.6276</td>
<td>0.6248</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.047</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>21.819</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>11.556</td>
<td>0.4029</td>
<td>0.9587</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>10.663</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>2.951</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>3.673</td>
<td>0.0026</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_7$</td>
<td>2.004</td>
<td>0.5294</td>
<td>0.1149</td>
</tr>
<tr>
<td>$\lambda_8$</td>
<td>2.160</td>
<td>0.6397</td>
<td>0.5403</td>
</tr>
<tr>
<td>$\lambda_9$</td>
<td>4.642</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_{10}$</td>
<td>149.620</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>$\lambda_{11}$</td>
<td>3.810</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_{12}$</td>
<td>16.540</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_{13}$</td>
<td>80.860</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda_{14}$</td>
<td>9.310</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
We provide details on the algorithm used here upon request.


Table 4
Parameter estimates (log models). This table reports the estimates for the structural parameters. The posterior mean is reported as the point estimate, posterior standard deviations and 5%–95% posterior intervals are reported in brackets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Models on log(VIX) with b = 0</th>
<th>Models on log(VIX) with b = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa )</td>
<td>0.014 (0.002)</td>
<td>0.019 (0.002)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>2.951 (0.064)</td>
<td>2.955 (0.063)</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.096 (0.001)</td>
<td>0.091 (0.001)</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>0.229 (0.044)</td>
<td>0.215 (0.045)</td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>0.111 (0.027)</td>
<td>0.084 (0.017)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.027 (0.005)</td>
<td>0.031 (0.005)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.002 (0.006)</td>
<td>0.074 (0.006)</td>
</tr>
<tr>
<td>( \eta )</td>
<td>0.078 (0.006)</td>
<td>0.074 (0.006)</td>
</tr>
</tbody>
</table>

The results presented in Section 5 gave no indication that including jumps necessarily improves the modelling of the VIX index. However, the SVV model differs substantially from previously studied specifications. For this reason we shall retain the possibility of jumps in the section and study whether the inclusion of a stochastic vol-of-vol factor alters our previous conclusions. Also note that the model presented here is affine, and contrary to the one-dimensional log model with \( b = 1 \), this implies that the VIX index is bounded by zero from below. Estimation of this model is, as before, by MCMC.\(^{16} \)

There are several motivations for considering this model. Firstly, the empirical results in the previous section motivate a more detailed study of the diffusion part of the process. Considering a stochastic volatility component is a natural extension for one-dimensional models and this approach has been successfully applied to other financial variables. Secondly, in the one-dimensional SDEs studied so far the jump probability is extremely high, so jumps cannot be interpreted as rare and extreme events, which is the main economic motivation for incorporating jumps into a diffusion model. The diffusion part is designed to create normal movements, whereas jumps contribute occasional shocks that are – because of their magnitude – unlikely to come from a pure diffusion process. If jumps were to occur very frequently these models may be poorly specified, or at least not compatible with their usual interpretation. A third motivation for considering the SVV specification is to capture the clustering in volatility of index changes that is evident from Fig. 1. As opposed to a transient shock, this feature is commonly modeled with a stochastic volatility component.\(^{16} \)

Table 5
Simulation results (log models). This table reports the p-values for all the statistics described in Section 3.

<table>
<thead>
<tr>
<th>Jump type</th>
<th>Data</th>
<th>( \log(VIX) ) with b = 0</th>
<th>( \log(VIX) ) with b = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>stdev</td>
<td>1.512</td>
<td>0.9250</td>
<td>0.9455</td>
</tr>
<tr>
<td>skew</td>
<td>0.427</td>
<td>0.9997</td>
<td>0.0110</td>
</tr>
<tr>
<td>kurt</td>
<td>21.819</td>
<td>0.9999</td>
<td>0.9880</td>
</tr>
<tr>
<td>avgmax10</td>
<td>11.556</td>
<td>0.9996</td>
<td>0.9804</td>
</tr>
<tr>
<td>avgmin10</td>
<td>10.663</td>
<td>0.0009</td>
<td>0.0002</td>
</tr>
<tr>
<td>perc1</td>
<td>3.673</td>
<td>0.2352</td>
<td>0.0664</td>
</tr>
<tr>
<td>perc5</td>
<td>2.904</td>
<td>0.6398</td>
<td>0.2017</td>
</tr>
<tr>
<td>perc95</td>
<td>2.160</td>
<td>0.6189</td>
<td>0.7076</td>
</tr>
<tr>
<td>perc99</td>
<td>4.642</td>
<td>0.9907</td>
<td>0.8342</td>
</tr>
<tr>
<td>absmax20</td>
<td>149.620</td>
<td>0.9997</td>
<td>0.9995</td>
</tr>
<tr>
<td>absmin20</td>
<td>3.810</td>
<td>0.0367</td>
<td>0.0545</td>
</tr>
<tr>
<td>max</td>
<td>16.540</td>
<td>0.9985</td>
<td>0.9309</td>
</tr>
<tr>
<td>min</td>
<td>17.360</td>
<td>0.0010</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

\(^{16} \)We provide details on the algorithm used here upon request.\n
6. Stochastic volatility of volatility

Having shown that the log-VIX models perform better than models for the VIX level, we now extend the log volatility specification a stochastic volatility-of-volatility (SVV) model as follows:

\[
    d \log(VIX_t) = \kappa (\theta - \log(VIX_t)) dt + \sqrt{\theta} dW_t + \Delta \log(N_t) dt, \\
    dV_t = \kappa \theta dW_t + \alpha \sqrt{V_t} dW'_t,
\]

where the correlation \( \rho \) between the two Brownian motions is assumed constant, but possibly non-zero. Considering a non-zero correlation case is essential in this set-up, as previous evidence points toward a strong dependence between the VIX and its volatility level. In addition to these test statistics, of no benefit. Interestingly, the jump models still struggle to capture the observed skewness of the VIX, but now the models tend to underestimate this statistic as the inclusion of jumps decreases the skewness in the models. Using \( b = 0 \) on the other hand, leads to significant misspecification.
volatility level, of between 21% and 22%. The correlation is, as expected, positive with high (and again virtually identical) estimates of 0.653 and 0.659.

The characteristics of the variance equation are very interesting, because this process differs somewhat from variance processes estimated from other financial variables. The speed of mean-reversion in the variance equation $\kappa_v$ is very high, at 0.11 for the diffusion model. This implies a very rapidly reverting process with an estimated value 10 times larger than for the VIX itself. Including a further jump component decreases this parameter only marginally, to a value of 0.097. The mean-reversion level for the variance $\theta_v$ is consistent with the estimate from the one-dimensional diffusion model. The estimate of 0.06 in the log volatility diffusion model reported in Table 4 is approximately equal to the average volatility level implied by our estimate for $\theta_v$. In order to visualize the variance $V$ over the sample period, we provide the estimated sample path of this latent variable in Fig. 2.

We have seen that including (state-independent) jumps into the SVV model changes parameter estimates only marginally, and this is probably because jumps occur only every six months, on average. Now, as desired, jump events concentrate only on exceptional outliers that cannot be explained with a more persistent stochastic vol-of-vol process. This is also reflected in the estimated jump sizes as, for all specifications, we obtain higher estimated jump sizes with a mean of 0.136 and a standard deviation of 0.103. This adds further evidence that jumps are now covering only the more extreme events. Also modelling negative jumps are of no major importance, as depicted by the estimated jump sizes depicted in Fig. 3.

These results pose an interesting question: Are jumps necessary at all once we account for stochastic vol-of-vol? To answer this consider the 5% percentile of the posterior distribution of $\lambda_0$, which is 0.003. This provides some statistical evidence in favor of including jumps, although they occur very infrequently. However, there is no evidence from our simulation results in Table 7 that including jumps improves the model. With or without jumps, the SVV model is capable of reproducing all the characteristics of the VIX that we consider. For both models it is the lower percentiles that are most difficult to reproduce, but still, the $p$-values for all models are between 0.05 and 0.95 so neither model can be rejected.

The rare occurrence of jumps is now similar to those found in the equity index market (Eraker et al., 2003). However, there is an important difference, because including jumps seems less important for volatility than it is for the index itself. The variance process of the VIX is much more quickly mean reverting and rapidly moving than the variance process of the S&P 500 index, omitting jumps from the specification has a lesser impact than it would have when variance is more persistent.

It is instructive to investigate the jumps in log-VIX events depicted in Fig. 3. Interestingly, the biggest estimated jump in the sample period is not obtained during the highly turbulent period of the banking crisis. This is because most of the movements are now captured by the stochastic vol-of-vol component. Instead there is an increased intensity

### Table 6
Parameter estimates (log vol-of-vol models). This table reports the estimates for the structural parameters. The posterior mean is reported as the point estimate, posterior standard deviations are given in parenthesis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Diffusion</th>
<th>Jump</th>
<th>Diffusion</th>
<th>Jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_v$</td>
<td>0.011</td>
<td>0.012</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>$\theta_v$</td>
<td>3.073</td>
<td>2.983</td>
<td>3.013</td>
<td>2.972</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>0.110</td>
<td>0.097</td>
<td>0.105</td>
<td>0.015</td>
</tr>
<tr>
<td>$\sigma_j$</td>
<td>0.349</td>
<td>0.330</td>
<td>0.016</td>
<td>0.035</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.183</td>
<td>0.162</td>
<td>0.183</td>
<td>0.162</td>
</tr>
<tr>
<td>$\lambda_0$</td>
<td>0.653</td>
<td>0.659</td>
<td>0.653</td>
<td>0.659</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.009</td>
<td>0.009</td>
<td>0.136</td>
<td>0.136</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Estimated variance paths. This figure depicts the estimated variance path (multiplied by 100) for the log(VIX) for the diffusion case.

**Fig. 3.** Estimated jumps. This figure depicts the average jump distribution for the model with normally distributed jumps.
of smaller jumps, so the vol-of-vol could adjust to capture even large outliers in the data. In other words, the clustering of large movements is difficult to create with a stochastic vol-of-vol component because its arrival was a total surprise and thus required a substantial upward jump.

Based on these observations we conclude that volatility jumps are re-sulting from relatively minor and very temporary corrections.\(^{17}\)

### 7. Applications to risk management

A standard task in risk management is to explore the effect of potential shocks in economic variables. The evolution of VIX can affect bank portfolios for many reasons, either indirectly as a measure of volatility, or more directly as the underlying of several derivative products such as futures, swaps and options. In this section, we take the most drastic scenario observed in our sample period and investigate the probability assigned to this scenario under different models for the VIX. To this end, we consider the evolution of VIX during the outburst of the banking crisis in autumn 2008, when VIX increased from 21.99 on September 2, to reach its all-time high of 80.06 on October 27. Preceding this peak, the index was increasing almost continually from the beginning of September, with only minor and very temporary corrections.\(^{17}\)

A possible strategy is to re-estimate the models using data until September 2008, as this would allow us to access the predictability of such a scenario. However, it is very unlikely that a pure statistical model based on our data could have predicted this scenario because since its inception the most extreme value of the VIX before October 2008 was 45.74, far away from the highs that were witnessed during the banking crisis. This is a deficiency of the data set, as even higher volatility levels were recorded during the global market crash of 1987, when the old volatility index VXO reached levels of more than 100%. For any risk management application it would be therefore crucial to take this pre-sample data into account, or to use parameter estimates from shocked data.

The question we address here is not the predictability of the banking crisis but whether the models, after observing such an extreme event (and incorporating it into the estimated parameters) are capable of generating such a scenario, or whether they still consider it impossible. Put differently, we ask how plausible is such a scenario under the different models, with parameters estimated after the event. For each model, we use the VIX value on September 2, 2008 (before the crisis) as our starting value and simulate the process until October 27, 2008 according to the parameter estimates presented in the two previous sections.\(^{18}\)

---

\(^{17}\) Another application of VIX index processes, to pricing VIX futures and options, has been removed for brevity. Details about these findings (which show that, consistent with Mencía and Sentana (in press), a stochastic vol-of-vol model has a marked impact on VIX derivatives) are provided upon request.

\(^{18}\) In addition, for the SVV models we use the estimated variance on September 2 as a starting value.
that they cover rare and extreme events. There is also strong statistical evidence in favor of time-varying jump intensities in these models. However since one-factor models are misspecified, it is likely that results for these models are distorted. Our simulation experiments show that the absolute benefit from the addition of jumps to one-factor models can be fairly small.

The only one-factor model that can explain a multitude of facets of the VIX is the non-affine log model. A yet more promising approach to capturing the extreme behavior of the VIX is the inclusion of a stochastic, mean-reverting variance process. This model passes all the specifications hurdles and yields superior results in our scenario analysis. It is also appealing because jumps are rare and extreme events, which only occur on days that can be linked to major political or financial news.

References


