



VIX Dynamics with Stochastic Volatility of Volatility

Andreas Kaeck

ICMA Centre, University of Reading

Carol Alexander

ICMA Centre, University of Reading

This Version: September 2010

ICMA Centre Discussion Papers in Finance DP2010-11

Copyright © 2010 Kaeck and Alexander. All rights reserved.

ICMA Centre • The University of Reading
Whiteknights • PO Box 242 • Reading RG6 6BA • UK
Tel: +44 (0)1183 788239 • Fax: +44 (0)1189 314741
Web: www.icmacentre.rdg.ac.uk

Director: Professor John Board, Chair in Finance

The ICMA Centre is supported by the International Capital Market Association



VIX Dynamics with Stochastic Volatility of Volatility

Andreas Kaeck Carol Alexander

ICMA Centre, Henley Business School, University of Reading, UK

This Version: September, 2010

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 dynamic 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.

JEL: C15, C32, G13, G15

Keywords: VIX, Volatility Indices, Jumps, Stochastic volatility of-volatility

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 (Carr and Wu, 2006, Jiang and Tian, 2007); their use in constructing trading strategies (Konstantinidi *et al.*, 2008) and for describing the dynamic behavior of equity return variance (Jones, 2003, Wu, 2010); and their information content regarding future volatility (Jiang and Tian, 2005), volatility and jump risk premia (Duan and Yeh, 2010), and the jump activity of equity returns (Becker *et al.*, 2009).

One of the most important strands of the literature focuses on the data generating process of the index itself. This is because a realistic model for volatility index dynamics is crucial for accurate pricing and hedging of volatility derivatives. 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. Whaley, 1993, Grunbichler and Longstaff, 1996 or Detemple and Osakwe, 2000, Psychoyios *et al.*, 2010).

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, Psychoyios, and Skiadopoulos (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 Psychoyios, Dotsis, and Markellos (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.¹ There is also evidence that volatility jumps constitute a relatively large fraction of the variability of volatility indices. Psychoyios, Dotsis, and Markellos (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 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. Here Eraker, Johannes, and Polson (2003) estimate about 1.5 volatility jumps per year when based on equity index data, yet Dotsis, Psychoyios, and Skiadopoulos (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 extends the empirical literature on continuous-time dynamics of volatility indices in several directions. 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. 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 are problematic because one loses the economic reasoning that jumps cover large, unexpected movements in the time-series. Hence, jump intensities of a very large order are likely to convey model misspecification. The estimation of non-affine models is, however, more difficult to handle, as discrete-time transition probabilities or characteristic

¹Dotsis, Psychoyios, and Skiadopoulos (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 not advisable as in this model volatility tends to either zero or infinity in the long run.

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 of volatility (*stochastic vol-of-vol* hereafter). This feature has, to our knowledge, not been studied for 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 an increased activity of the jump part. Our results are interesting because this distinction allows for two separate categories of jumps: transient (unexpected) jumps and jumps due to persistent high volatility regimes. We find that considering such an extension is of first-order importance and that the estimated variance process for VIX is extremely erratic and mean-reverts very quickly. We further investigate whether both jumps and stochastic vol-of-vol are necessary but our results regarding this issue are mixed.

Fourthly, we provide extensive simulation results that allow us to gage the absolute performance of all models under consideration. We use 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 several non-statistical metrics. In particular we perform a scenario analysis exercise and study the impact of the models on the pricing of simple derivatives.

We proceed as follows: Section I introduces the affine and non-affine one-factor models used; Section II describes our econometric estimation methodology. Section III provides details on the data set. In Section IV we provide estimation results for various alternative one-factor processes. Section V introduces and presents results for the stochastic vol-of-vol model. We provide risk management and derivatives pricing applications in Section VI and Section VII concludes.

I Models

Most models proposed for describing volatility or variance dynamics agree on its mean-reverting nature.² 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.³

²Only 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.

³This is however contradicted by the empirical findings in Dotsis, Psychoyios, and Skiadopoulos (2007) who report that a Merton-type return model for volatility outperforms the mean-reverting specifications in their sample. Yet, the authors do not consider mean reverting log-volatility processes and thus it is not entirely

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, Dotsis, Psychoyios, and Skiadopoulos (2007) rely on the square-root and a Merton-type jump model for volatility and Psychoyios, Dotsis, and Markellos (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 has been successfully applied in other contexts, such as option pricing or spot index modeling (see Christoffersen, Jacobs, and Mimouni (2010), Chernov, Gallant, Ghysels, and Tauchen (2003)). Especially for option pricing applications researchers often favor square-root specifications, as they retain 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, Johannes, and Polson (2003) (using return data) and Broadie, Chernov, and Johannes (2007) (using both return and option data) find severe misspecifications when jumps in volatility are omitted and document the outperformance of variance specifications with exponential upward jumps. Dotsis, Psychoyios, and Skiadopoulos (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 Psychoyios *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 V. First we study models that are nested in the following specification:

$$dX_t = \kappa(\theta - X_t) dt + \sigma X_t^b dW_t + Z_t dJ_t \quad (1)$$

where X either denotes the value of the volatility index or its logarithm, κ is the speed of mean reversion, θ determines the long term value of the process and σ is a constant in the diffusion term. The exponent b 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, Gallant, Ghysels, and Tauchen (2003), this is a very mild restriction for yearly volatility.⁴

In terms of jump distributions we assume that J is a Poisson process with time varying intensity $\lambda_0 + \lambda_1 X_{t-}$.⁵ 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 (and to the jumps in default intensity models). 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 η_J , 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 μ_J and

clear whether their result is driven by the fact that log specification outperforms other models or whether the result is due to the underperformance of mean-reverting processes.

⁴To avoid this one could also model not the VIX directly, but its value minus this lower bound.

⁵We use the standard shorthand notation X_{t-} for the left limit, hence $X_{t-} \equiv \lim_{s \uparrow t} X_s$

standard deviation σ_J because the support of this model is not restricted to positive numbers and the log volatility may become negative.⁶

II Econometric Methodology

A 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, Psychoyios, and Skiadopoulos (2007) use conditional maximum likelihood methods to estimate the structural parameters of several alternative processes for six different volatility indices. Psychoyios, Dotsis, and Markellos (2010) apply 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

$$p(\Theta, \mathbf{Z}, \mathbf{J} | \mathbf{X}) \propto p(\mathbf{X} | \Theta, \mathbf{Z}, \mathbf{J}) p(\Theta, \mathbf{Z}, \mathbf{J}).$$

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

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 uninformative, hence our parameter estimates are driven by the information in the data and not the prior. More details about these distributions are given in the appendix. 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

⁶In fact, we have also estimated all models with both normally and exponentially distributed jump size, so that we may gauge 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.

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

two time steps the process evolves according to

$$X_{t+1} = X_t + \sum_{i=0}^{1/h-1} \left[\kappa (\theta - X_{t+ih}) h + \sqrt{h} \sigma X_{t+ih}^b \varepsilon_{t+(i+1)h} + Z_{t+(i+1)h} J_{t+(i+1)h} \right]$$

where h denotes the discretization step, ε_t denote standard normal variates and the jump process is discretized by assuming that the event $J_{t+h} = 1$ occurs with probability $h(\lambda_0 + \lambda_1 X_t)$. This approximation converges (under some regularity conditions) to the true continuous-time process as h approaches zero. Therefore choosing h to be small should lead to a negligible discretization bias. But in reality the frequency of the observed data cannot be determined 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.⁸

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 ($1/h - 1$ times the sample size). 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 $h = 0.25$. Jones (2003) reports that, for equity index data, taking h to be of this order reduces the discretization bias noticeably.

The posterior for our parameter estimation therefore includes the augmentation of \mathbf{X} by unobservable high-frequency observations \mathbf{X}^u and yields

$$p(\Theta, \mathbf{Z}, \mathbf{J}, \mathbf{X}^u | \mathbf{X}) \propto p(\mathbf{X}, \mathbf{X}^u | \Theta, \mathbf{Z}, \mathbf{J}) p(\Theta, \mathbf{Z}, \mathbf{J}).$$

Note that although we generate a distribution of each augmented data point, we have no interest in the density of \mathbf{X}^u 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 and 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:

⁸The 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, Psychoyios, and Skiadopoulos (2007) volatility indices can jump far too frequently for this to be negligible. However, if the jump intensity is much lower, as in Eraker, Johannes, and Polson (2003), a daily discretization does not introduce any discernible error.

1. Draw the latent variables:	$p \left(\mathbf{X}^{u(g)} \mid \Theta^{(g-1)}, \mathbf{Z}^{(g-1)}, \mathbf{J}^{(g-1)}, \mathbf{X} \right)$ $p \left(\mathbf{Z}^{(g)} \mid \mathbf{X}^{u(g)}, \Theta^{(g-1)}, \mathbf{J}^{(g-1)}, \mathbf{X} \right)$ $p \left(\mathbf{J}^{(g)} \mid \Theta^{(g-1)}, \mathbf{Z}^{(g)}, \mathbf{X}^{u(g)}, \mathbf{X} \right)$
2. Draw structural parameters:	$p \left(\Theta^{(g)} \mid \mathbf{X}^{u(g)}, \mathbf{Z}^{(g)}, \mathbf{J}^{(g)}, \mathbf{X} \right)$

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.⁹ More details about the exact distributions and algorithms for our case are given in the appendix.

B 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.¹⁰

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 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 (*stdev*), 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 linked to jump behavior of the process. We use the highest positive and negative jump in the index (*minjump* and *maxjump*), the average over the 10 largest positive jumps (*avgmax10*) and the average over the 10 largest negative jumps (*avgmin10*). These statistics shed light on whether the model can replicate the observed outliers.
- In order to investigate the clustering of the outliers we use the month (20 trading days)

⁹A standard reference including a wide range of Metropolis algorithms is Robert and Casella (2004).

¹⁰For 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).

¹¹Very similar statistics have been used recently for equity index dynamics in Kaeck and Alexander (2010).

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 will struggle 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 the parameter uncertainty into account because it draws the structural parameters randomly from the posterior density.

III Data

The VIX volatility index is constructed from all standard European S&P 500 index options for the two delivery dates straddling 30 days to maturity. These are used to infer a constant 30-days-to-maturity volatility estimate. CBOE publishes this index on a daily basis and makes it publicly available on their website (www.cboe.com). The construction methodology is based on the results in Britten-Jones and Neuberger (2000) and hence it allows one to regard VIX an estimate of volatility that is *model free* under some fairly unrestrictive assumptions on the equity index data generation process. We use daily time series data from January 1990 until May 2010 for the VIX index, which is the longest time series available at the time of writing.

VIX and its logarithm are depicted in Figure 1. As expected, all high volatility periods coincide with either major political events or financial 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

Figure 1: **VIX Index.** This figure depicts both VIX and its log, as well changes in their values. The sample period for the VIX from January 1990 until end of May 2010.

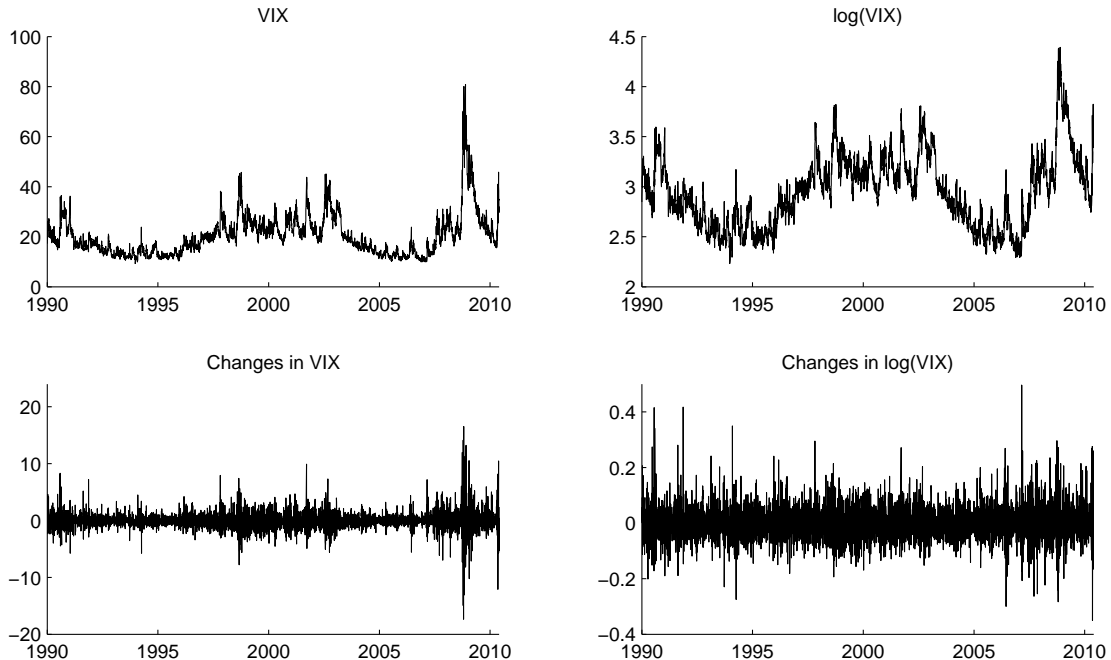


Table 1: **Descriptive Statistics.** This table reports sample statistics for levels and first differences of the VIX. The sample period for the VIX is from January 1990 until May 2010.

	mean	std dev	skewness	kurtosis	min	max
Level	20.32				9.31	80.86
First Difference	0.003	1.512	0.427	21.819	-17.36	16.54

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.

IV Estimation Results

A 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

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.

	Models on VIX with $b = 0.5$			Models on VIX with $b = 1$		
κ	0.016 (0.003) [0.011, 0.02]	0.037 (0.003) [0.033, 0.041]	0.051 (0.003) [0.047, 0.055]	0.014 (0.003) [0.01, 0.019]	0.029 (0.003) [0.024, 0.034]	0.039 (0.003) [0.034, 0.044]
θ	20.496 (1.187) [18.667, 22.544]	13.747 (0.567) [12.806, 14.665]	12.311 (0.473) [11.525, 13.076]	20.510 (1.357) [18.566, 22.951]	13.294 (0.561) [12.358, 14.207]	13.033 (0.427) [12.337, 13.739]
σ	0.289 (0.004) [0.283, 0.296]	0.229 (0.003) [0.223, 0.234]	0.214 (0.003) [0.209, 0.219]	0.062 (0.001) [0.06, 0.063]	0.050 (0.001) [0.049, 0.051]	0.048 (0.001) [0.047, 0.049]
λ_0		0.107 (0.016) [0.082, 0.134]			0.082 (0.012) [0.063, 0.102]	
λ_1			0.013 (0.002) [0.01, 0.016]			0.007 (0.001) [0.005, 0.008]
η_J		2.376 (0.185) [2.104, 2.706]	1.584 (0.102) [1.428, 1.759]		2.708 (0.21) [2.399, 3.074]	2.299 (0.161) [2.057, 2.584]

column, we estimate a speed of mean reversion κ 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 θ is about 20.5% which is close to the unconditional mean of the process in Table 1. Our parameter estimate for σ is 0.289.¹²

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 of the level of the VIX. We have also estimated all models with λ_0 and λ_1 being simultaneously different from zero.¹³ 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 κ estimate. Furthermore, in the jump models the estimates for the second drift parameter θ drop to about 12-14%, a result that is expected because θ carries a different interpretation once jumps are included. To obtain the long-term volatility we have to adjust θ 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 σ decreases in all jump models since part of the variation in the VIX is now

¹²Note that this model was previously studied in Dotsis, Psychoyios, and Skiadopoulos (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 Dotsis, Psychoyios, and Skiadopoulos (2007) imply a 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 24.54%.

¹³These results are omitted for expositional clarity, but they are available from the authors upon request. The parameter estimates for these models reveal that jump probabilities are mainly driven by the state-dependent jump part as λ_0 is close to zero. Therefore, the evidence appears to point toward state-dependent jumps. We return to this observation later on.

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%.¹⁴ As we estimate more jumps in this case, the average jump size decreases to only 1.58 VIX points.

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 κ 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 rapid 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 σ , 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.

B Jump-Diffusion Models on the Log of the VIX Level

Structural parameter estimates for the log-VIX models are reported in Table 4. We consider the models with $b = 0$ first, shown in the left side of the table. The mean reversion speed κ is more consistent across models with and without jumps, taking values between 0.014 and 0.017. The long-term level θ for the log process is estimated to be 2.951 in the pure diffusion model, a value that implies a long-term volatility level of about 19%. The value for this parameter is again dependent on the estimated jump parameters and hence it drops in the jump models. The implied long-term volatility level however hardly changes, for example

¹⁴This estimate is based on an average VIX level of about 20%.

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

Jump Distribution Jump Type	Data	VIX with $b = 0.5$			VIX with $b = 1$		
		no	exp		no	exp	
			λ_0	λ_1		λ_0	λ_1
stadev	1.512	0.9988	0.6276	0.6248	0.8644	0.4395	0.1993
skew	0.427	1.0000	0.0000	0.0000	0.9991	0.0000	0.0000
kurt	21.819	1.0000	0.8474	0.9985	0.9972	0.6537	0.9048
avgmax10	11.556	1.0000	0.4029	0.9587	0.9916	0.1721	0.2386
avgmin10	-10.663	0.0000	0.0000	0.0000	0.0153	0.0000	0.0000
perc1	-3.673	0.0026	0.0000	0.0000	0.3545	0.0045	0.1162
perc5	-2.004	0.9294	0.1149	0.3068	0.7652	0.4198	0.6821
perc95	2.160	0.6397	0.5043	0.0050	0.5909	0.5245	0.1058
perc99	4.642	1.0000	0.0051	0.0108	0.9687	0.0093	0.0008
absmax20	149.620	1.0000	1.0000	1.0000	0.9922	1.0000	0.9999
absmin20	3.810	0.0000	0.0000	0.0000	0.0016	0.0009	0.0018
maxjump	16.540	1.0000	0.5571	0.9122	0.9884	0.3621	0.4735
minjump	-17.360	0.0000	0.0000	0.0000	0.0110	0.0000	0.0001
max	80.860	0.9984	1.0000	0.9998	0.8521	0.9968	0.9460
min	9.310	0.9998	0.9514	0.9592	0.9906	0.7969	0.7724

our results in the state-independent and exponential jump model implies a similar long-term volatility level of 19.9%. Estimates for σ vary across models, between 0.04 and 0.06. The jump likelihood in the log volatility model is again very high, with daily jump probabilities of 20% or more, which implies more than 50 jumps per year. The average jump probability for the time-varying jump intensity model is of larger magnitude. The normally distributed jump sizes have mean 0.03 with a standard deviation of around 0.08. Parameter estimates for the log model with additional dependence of the diffusion term on the level of the VIX are reported in the right half of Table 4. The only noteworthy feature of our estimates here is that jump sizes are higher, with an estimated mean of about 0.08 for both models.

Simulation results for the log models are presented in Table 5. Models with $b = 1$ perform quite well in producing samples with similar characteristics as the observed VIX time series. The only characteristic that can be rejected at a 5% significance level is the skewness. The observed statistic is 0.427, but the simulations imply a smaller statistic in 97.96% of the cases. Apart from this, the pure diffusion model produces realistic samples. This is true in particular of the large jumps in the VIX. For example the large negative and positive jumps of more than -17 and 16 VIX points respectively, creates no obstacle for the model. Including jumps into the processes can improve some of the statistics we use, but overall the inclusion of jumps is, at least as far as these test statistics are concerned, 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.

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.

	Models on log(VIX) with $b = 0$			Models on log(VIX) with $b = 1$		
κ	0.014 (0.002) [0.01, 0.018]	0.016 (0.002) [0.012, 0.019]	0.017 (0.002) [0.013, 0.02]	0.014 (0.002) [0.01, 0.018]	0.019 (0.002) [0.015, 0.023]	0.022 (0.002) [0.018, 0.026]
θ	2.951 (0.064) [2.848, 3.055]	2.562 (0.076) [2.431, 2.673]	2.534 (0.077) [2.403, 2.649]	2.955 (0.063) [2.858, 3.064]	2.098 (0.115) [1.898, 2.274]	2.152 (0.096) [1.988, 2.298]
σ	0.060 (0.001) [0.059, 0.062]	0.043 (0.001) [0.041, 0.045]	0.040 (0.002) [0.037, 0.044]	0.020 (0) [0.02, 0.021]	0.016 (0) [0.016, 0.017]	0.016 (0) [0.015, 0.016]
λ_0		0.229 (0.044) [0.164, 0.303]			0.215 (0.045) [0.154, 0.296]	
λ_1			0.111 (0.027) [0.069, 0.156]			0.084 (0.017) [0.06, 0.114]
μ_J		0.027 (0.005) [0.019, 0.036]	0.022 (0.005) [0.015, 0.031]			
σ_J		0.082 (0.006) [0.074, 0.092]	0.074 (0.007) [0.065, 0.085]			
η_J					0.078 (0.006) [0.068, 0.088]	0.074 (0.006) [0.065, 0.084]

V 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 to the following stochastic volatility-of-volatility (SVV) model:

$$\begin{aligned}
 d \log(\text{VIX}_t) &= \kappa [\theta - \log(\text{VIX}_t)] dt + \sqrt{V_t} dW_t + Z_t dJ_t \\
 dV_t &= \kappa_v (\theta_v - V_t) dt + \sigma_v \sqrt{V_t} dW_t^v
 \end{aligned}$$

where the correlation ρ 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 a stochastically moving volatility, we allow for normally distributed jumps as before. Estimation of this model is by MCMC, as before, and we describe the exact algorithm in the appendix.

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*

Table 5: Simulation Results (Log models). This table reports the p -values for all the statistics described in Section II.

Jump Distribution Jump Type	Data	log(VIX) with $b = 0$			log(VIX) with $b = 1$		
		no	normal		no	exp	
			λ_0	λ_1		λ_0	λ_1
stadev	1.512	0.9250	0.9455	0.8545	0.6069	0.1484	0.0895
skew	0.427	0.9997	0.0110	0.0257	0.9796	0.0001	0.0004
kurt	21.819	0.9999	0.9880	0.9764	0.9426	0.5219	0.4089
avgmax10	11.556	0.9996	0.9804	0.9450	0.8881	0.1114	0.0627
avgmin10	-10.663	0.0009	0.0002	0.0048	0.1601	0.0747	0.2679
perc1	-3.673	0.2352	0.0664	0.2159	0.6917	0.6236	0.8047
perc5	-2.004	0.6930	0.2017	0.3245	0.7873	0.7829	0.8531
perc95	2.160	0.6189	0.7076	0.5310	0.4228	0.0554	0.0444
perc99	4.642	0.9907	0.8342	0.7257	0.7814	0.0165	0.0133
absmax20	149.620	0.9997	0.9995	0.9951	0.8903	0.8846	0.6811
absmin20	3.810	0.0367	0.0545	0.0893	0.0906	0.1295	0.2282
maxjump	16.540	0.9985	0.9309	0.8881	0.8572	0.1603	0.0953
minjump	-17.360	0.0010	0.0016	0.0096	0.1253	0.0633	0.2028
max	80.860	0.9452	0.9635	0.8961	0.6763	0.6506	0.4037
min	9.310	0.9929	0.9770	0.9742	0.9466	0.9315	0.9539

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.

	κ	θ	κ_v	$\theta_v \times 100$	$\sigma_v \times 10$	ρ	λ_0	μ_J	σ_J
mean	0.011	3.073	0.110	0.349	0.183	0.653			
standard dev	(0.002)	(0.086)	(0.013)	(0.016)	(0.01)	(0.038)			
mean	0.012	2.983	0.097	0.330	0.162	0.659	0.009	0.136	0.103
standard dev	(0.002)	(0.072)	(0.015)	(0.035)	(0.016)	(0.039)	(0.004)	(0.054)	(0.027)

movements, whereas jumps contribute occasional shocks that are – because of their magnitude – unlikely to come from a pure diffusion process.¹⁵ 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 Figure 1. As opposed to a transient shock, this feature is commonly modeled with a stochastic volatility component.

Table 6 reports the estimated parameters of the SVV model, first without jumps and then with normally distributed jumps. The speed of mean reversion parameter κ is lower than in any previously reported model, with an estimate of 0.011 and 0.012. As mentioned above, κ is likely to be distorted upward when a model cannot capture large negative outliers, so this result indicates that SVV models are more consistent with large downward moves than models without stochastic volatility. Furthermore, both models imply a long-term volatility

¹⁵Eraker, Johannes, and Polson (2003) argue that this justifies a more restrictive prior on the jump likelihood parameters.

Figure 2: **Estimated Variance Paths.** This figure depicts the estimated variance path (multiplied by 100) for the log(VIX) for the diffusion case.

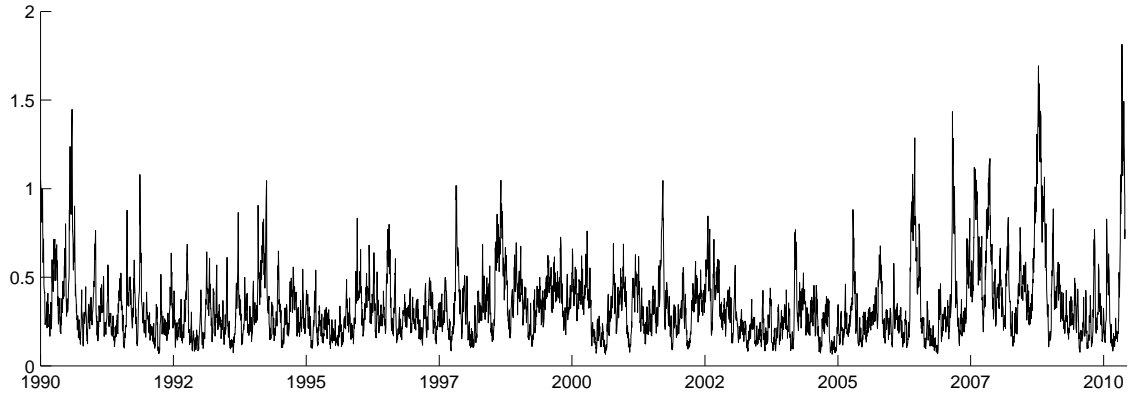
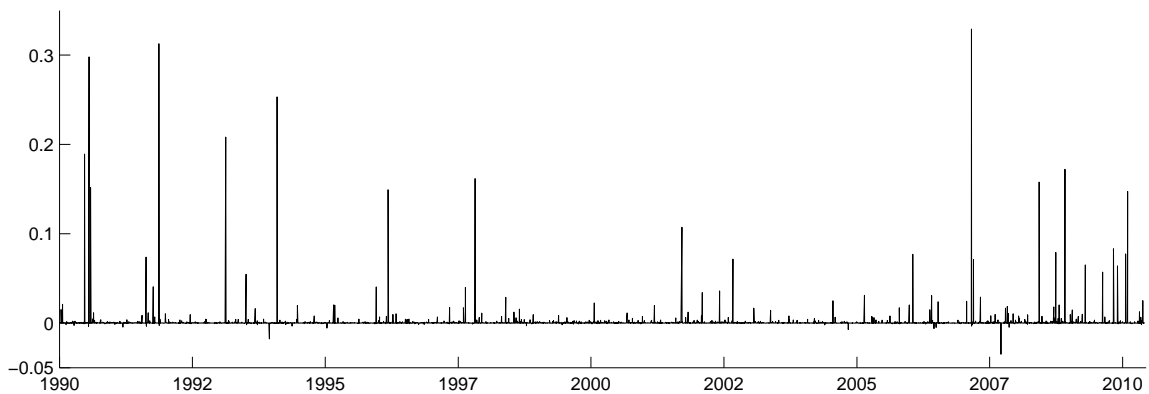


Figure 3: **Estimated Jumps.** This figure depicts the average jump distribution for the model with normally distributed jumps.



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 κ_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 θ_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 θ_v . In order to visualize the variance V over the sample period, we provide the estimated sample path of this latent variable in Figure 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

Table 7: **Simulation Results (Stochastic vol-of-vol models).**

Stochastic-vol-of-vol Models on log(VIX)			
	Data	<i>p</i> -values	
Jump Distribution		no	normal
stadev	1.512	0.260	0.245
skew	0.427	0.713	0.338
kurt	21.819	0.829	0.791
avgmax10	11.556	0.607	0.505
avgmin10	-10.663	0.382	0.342
perc1	-3.673	0.924	0.911
perc5	-2.004	0.922	0.913
perc95	2.160	0.129	0.133
perc99	4.642	0.353	0.289
absmax20	149.620	0.728	0.748
absmin20	3.810	0.294	0.308
maxjump	16.540	0.563	0.457
minjump	-17.360	0.310	0.275
max	80.860	0.369	0.381
min	9.310	0.777	0.762

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 negative jumps are of no major importance, as depicted by the mean jump sizes depicted in Figure 3. Although negative jumps are clearly a feature of volatility indices, their occurrence is rather a correction of previous large positive jumps. Indeed our results indicate no negative jump of significant size at all, over the entire sample period.

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 λ_0 , which is 0.003. This provides some statistical evidence in favor of including jumps, although they occur very infrequently. However, there is not 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 Figure 3. Interestingly, the biggest estimated jump in the sample period are 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 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 was best captured with a stochastic vol-of-vol component. One of the largest jumps in our sample was in November 1991, when the VIX jumped from less than 14 to over 21 in one day. This jump was preceded by several tranquil months with little movements in the VIX. The same applies to the jumps in February 1993 and in February 1994. Another large jump is estimated in February 2007. Prior to this, volatility was bounded between about 10 and 13 percent for many months. Then a slump in the Chinese stock market created a knock-on effect for Europe, Asia and North America with substantial losses for all major equity indices on 27 February. This left financial markets in doubt over economic prospects, and the VIX jumped up by more than 7 points. This jump is difficult to create with a stochastic vol-of-vol component because its arrival came as a total surprise and thus required a substantial upward jump. Based on these observations we conclude that volatility jumps are required, but only for surprising events triggered by totally unexpected political or financial news. Note also, that the jumps estimated by the model occur during periods of low VIX levels thus there is no evidence in this model that suggests that jumps are more likely when VIX levels are high.

VI Applications to Risk Management and Derivative Pricing

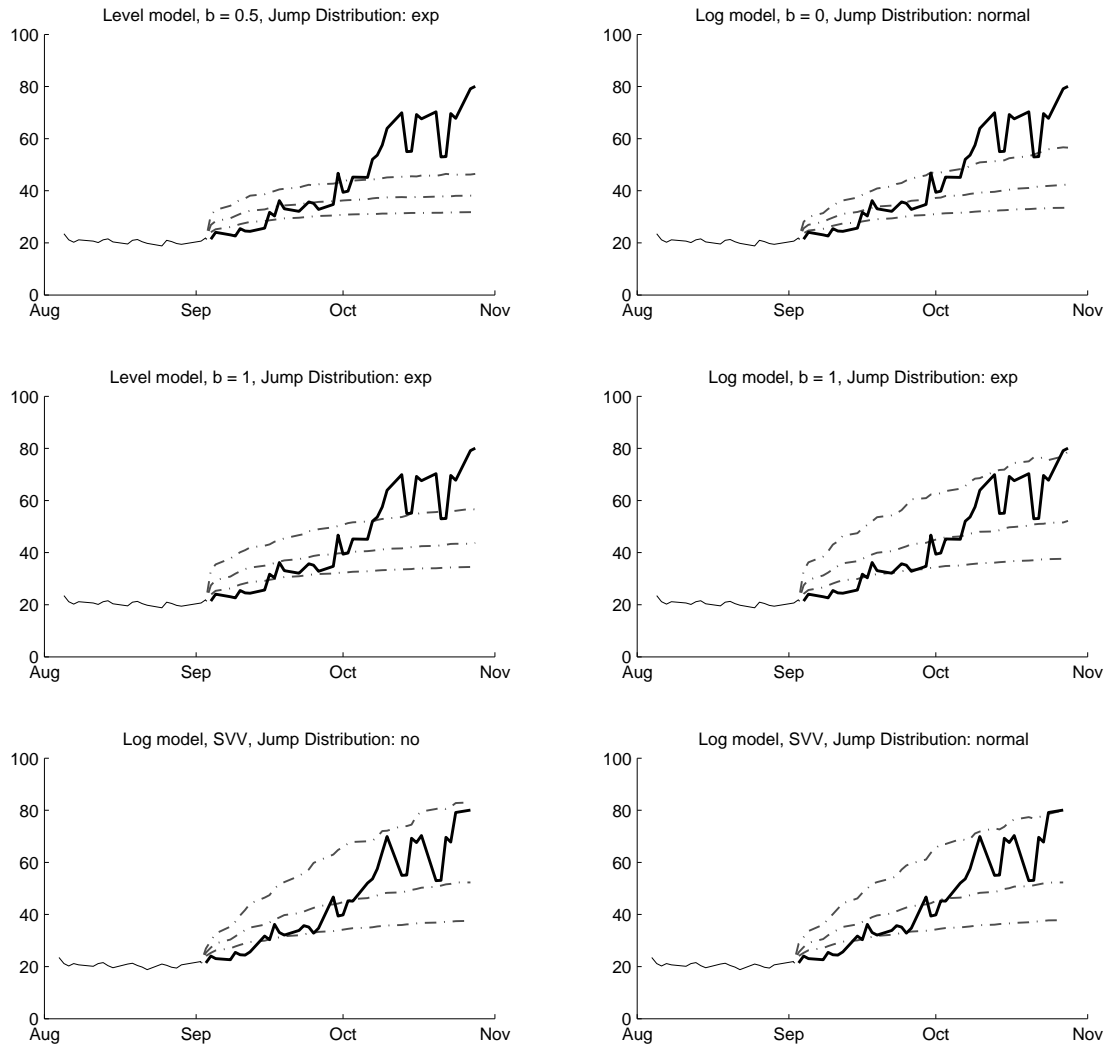
So far we have judged alternative models for the VIX purely on econometric grounds, so in this section we provide two applications to standard risk management and derivative pricing. We begin by analyzing differences between the scenarios that are generated by alternative models and then we consider some implications for pricing derivatives on the VIX index.

A Scenario Analysis

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 only few weeks later on October 27. Preceding this peak, the index was increasing almost continually from the beginning of September, with only minor and very temporary corrections.

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

Figure 4: **Simulated VIX 2008.** This figure depicts the true evolution of the VIX during the beginning of the banking crisis in 2008. In addition, we plot 95%, 99% and 99.9% percentiles.



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 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.¹⁶ Then, after simulating 100,000 paths, we gauge the likely range of values produced by the models by calculating percentiles for the two-month period. In each simulation the parameter values are drawn randomly from the posterior distribution, so that the analysis takes account of the uncertainty in estimated parameters.

Figure 4 illustrates the results of this exercise for six of the models. For the one-factor models we consider the most general specifications, with time-varying jump probabilities. Other assumptions on the jump part of the processes lead to virtually identical conclusions and so we omit these for expositional clarity. In both affine and non-affine models of the VIX itself the index ends up far beyond the 99.9%-percentile. Log models fare better but still assign only a tiny probability to the likelihood of the observed path. The best among the one-factor models is the log model with additional dependence of the diffusion coefficient on the VIX level. This finding confirms our previous evidence that such a modeling approach yields the most realistic results, among all the one-factor models considered. SVV models also do a good job, as for both processes the actual time series ends between the 99% and 99.9% percentiles. Indeed, given that our sample consists of almost 150 such two months periods, we would hope that such a one-off scenario is predicted in less than 1% of the cases. We conclude that only the one-factor log model with $b = 1$ and the stochastic volatility-of-volatility models provide accurate assessments of the likelihood of the banking crisis scenario.

B Implications for Derivatives Pricing

To study the effect of incorporating stochastic volatility in the VIX process on standard derivatives, such as VIX futures and variance swaps, we compare the empirically observed term structures for VIX futures with the term structures generated by the different models. We focus on three different models: both log and level non-affine one-factor models and the SVV model augmented with jumps.¹⁷ The modeling of VIX futures requires parameters from the risk-neutral probability measure rather than real-world parameter estimates, so we follow Dotsis, Psychoyios, and Skiadopoulos (2007) and assume that the market prices of diffusive and jump risk are zero.¹⁸

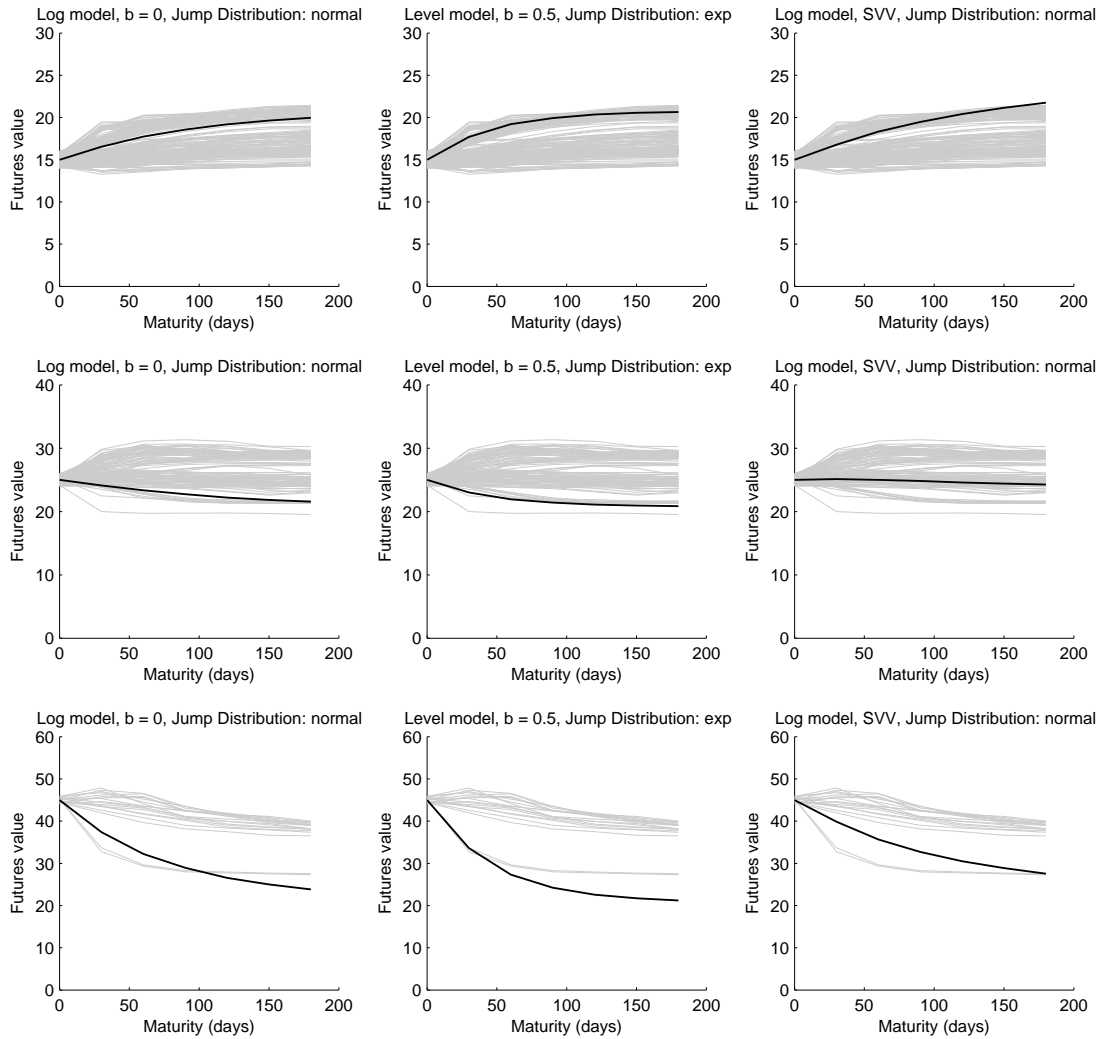
The prices of VIX futures with maturity between one and six months are downloaded from the CBOE website and are available from March 2004 until the end of the sample. We assign a term structure on a given day to one of three volatility buckets (15%, 25% and 45%) if the VIX was within $\pm 1\%$ of these target values. This procedure provides 189 and 107 term structures for low and medium volatility levels respectively and 18 term structures for the 45% VIX level. The collected data illustrate the different shapes of volatility term structures that are empirically observable. We then calculate the term structure for the three alternative models using Monte-Carlo simulation. For the one-factor models the term-structure is fixed

¹⁶In addition, for the SVV models we use the estimated variance on September 2 as a starting value.

¹⁷Again, we do this to focus on the main results but we have performed the same exercise for all the models considered in this paper.

¹⁸Alternatively, we could shift some of the parameters by an additive or multiplicative (yet arbitrary) constant as in Johannes (2004), but since our derivatives pricing exercise is mainly relative in nature and because the effect of such a shift on the different models is difficult to gauge, we prefer to set risk premia to zero.

Figure 5: **VIX Futures Term Structure.** This figure depicts the observed term structures between 2006 and 2010. In black we depict the term structure implied by the models and in gray the empirically observed term structures between March 2004 and May 2010.



using the parameter estimates, however for the SVV models the term structures also depend on the spot level of the VIX variance V_t at initiation of the contract. In our simulation we set this value to the mean of the estimated variance on the days in the sample where VIX is $\pm 1\%$ of the target VIX value.

Figure 5 plots the observed VIX futures term structures in gray and the theoretical term structure from each model is superimposed in black. For the one-factor models, the theoretical term structure remains with the empirically observed data for low and medium volatility levels, but for high volatility levels both models imply a rapidly declining term structure which is incompatible with the futures prices, especially long-term futures prices. Most of the

Table 8: **VIX Option Pricing.** This table reports call option prices for the different models. All option prices are calculated by Monte-Carlo simulation.

VIX level	Moneyness: 1.2			Moneyness: 1.5		
	15%	25%	45%	15%	25%	45%
Days to Maturity: 30						
Level model, $b = 0.5$, Jump Dist: exp	1.72	0.42	0.03	0.52	0.05	0.00
Log model, $b = 0$, Jump Dist: normal	0.98	0.73	0.47	0.23	0.15	0.08
Level model, $b = 1$, Jump Dist: exp	1.67	0.76	0.28	0.60	0.17	0.03
Log model, $b = 1$, Jump Dist: exp	1.40	1.35	1.43	0.51	0.50	0.55
Log model, SVV, Jump Dist: no	1.19	1.12	1.24	0.41	0.37	0.44
Log model, SVV, Jump Dist: normal	1.21	1.18	1.26	0.41	0.40	0.43

observed term structures at this volatility level decrease more slowly than the models imply. Only for the stochastic vol-of-vol model all theoretical prices remain within the empirically observed range. However, overall the term structures observed in the market point toward a weakness of all the models we consider. Since the VIX futures price converges at long maturities toward the estimated long-term VIX level, there is little variation in the shapes produced by the models. A possible remedy for this is to consider a further stochastic factor that drives the long-term volatility level. This way, the shape of the term structure will also depend on the level and the dynamics of this second factor. If this factor is positively related to the VIX level itself, it is likely to generate more realistic term structures.

The impact of the data generating process on VIX option prices is investigated in Table 8. We use standard European call options on the VIX with 30 days to maturity. Since the difference between option pricing models is most visible in out-of-the-money option prices we study options with two different moneyness (defined as strike divided by spot value) categories, 120% and 150%.¹⁹ These results are in line with our previous findings. The one-factor log model with $b = 1$ yields prices relatively close to the prices calculated for the SVV specifications. However, the difference between these and other one-factor models can be substantial. For example the affine level model assigns almost no value to far out-of-the-money (OTM) options, whereas the prices in the SVV models are still considerable. It is interesting that prices for OTM calls in the SVV models are relatively insensitive to the current VIX level. This highlights the important effect that the spot variance may have on VIX option prices. When the VIX is at low levels, its variance tends to be low as well. But in high volatility regimes, although the mean-reversion pulls the process back toward a smaller value, the high variance of the process leads to option prices that are comparable to those at lower VIX levels. This effect is also visible for the one-factor log model with $b = 1$, but in other one-factor models the option prices are far too low because drift term dominates even in high volatility regimes.

¹⁹Again, we have performed the same exercise for other maturities and moneyness levels, but for brevity concentrate on the most important short term out-of-the money call options here.

VII Conclusion

This paper has studied alternative jump-diffusion models for the VIX volatility index, considering two broad modeling approaches, i.e. to model the VIX directly or its log value. Our models include one-factor affine and non-affine diffusion and jump-diffusion models, and two-factor stochastic volatility models. We evaluate these models using probability values for a wide range of statistics and assess their performance for some risk management and derivatives pricing applications.

As in Dotsis, Psychoyios, and Skiadopoulos (2007) we find that modeling the VIX log returns (equivalently, the log value of VIX) is superior to modeling its level. Beyond this we present a variety of novel contributions to the literature. First, we find that non-affine models, in which the diffusion term is proportional to the VIX level or log respectively, are far superior to their affine counterparts. The main reason for this is that non-affine models accommodate a more rapidly moving VIX during high volatility regimes. Not only are affine models unable to reproduce the observed characteristics of the VIX, they also assign too great an intensity to the jump processes. This is problematic, since the intuition of introducing jumps is 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 specification 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. For modeling the VIX futures term structure a stochastic mean reversion factor appears to be important. Finally we show that VIX option prices that are generated by different models can vary significantly. This emphasizes the importance of our research, as the model risk involved in choosing a process for the VIX can be substantial.

Appendix

This appendix provides distributions and algorithms used in the MCMC estimation. General references for Bayesian statistics and MCMC methods are Geweke (2005) and Robert and Casella (2004) to which we refer for details on the Gibbs sampler and Metropolis steps.

A Complete Conditionals and Prior Information - One Factor Models

We first describe the draws of the augmented data set. We index the observations of the VIX (or its logarithm) from 0 to T with a step size of one. This way, real observations have integer-valued indices. Other indices denote unobserved data. In this notation, \mathbf{X}^u consists of all X_{ih} with non-integer ih , $i = 0, \dots, T/h$, and \mathbf{X} collects remaining X_{ih} (the *true observations*). In order to update X_{ih} for non-integer ih , its posterior is given (up to proportionality) by

$$p \left(X_{ih} \mid X_{(i-1)h}, \mathbf{Z}, \mathbf{J}, \Theta \right) \times p \left(X_{(i+1)h} \mid X_{ih}, \mathbf{Z}, \mathbf{J}, \Theta \right)$$

where $p(\cdot)$ is a general notation for a probability density. Since the product of two normal densities is non-standard, we use a Metropolis step. We propose a value from the first of the two distributions and accept it based on the likelihood ratio of the second density. For more details on this we refer to Jones (1998).

We turn to the latent state variables \mathbf{Z} and \mathbf{J} , and the parameter vector Θ . We choose prior distributions and hyper-parameters as listed below. We underline for prior and overline for posterior distributions to simplify the notation.

- \mathbf{J} : The posterior distribution of J_{ih} is Bernoulli with jump probability of $p = A/(A+B)$, where

$$A = h(\lambda_0 + \lambda_1 X_{(i-1)h}) \exp \left\{ -\frac{\left(X_{ih} - X_{(i-1)h} - \kappa(\theta - X_{(i-1)h})h - Z_{ih} \right)^2}{2h\sigma^2 X_{(i-1)h}^{2b}} \right\},$$

$$B = (1 - h(\lambda_0 + \lambda_1 X_{(i-1)h})) \exp \left\{ -\frac{\left(X_{ih} - X_{(i-1)h} - \kappa(\theta - X_{(i-1)h})h \right)^2}{2h\sigma^2 X_{(i-1)h}^{2b}} \right\}.$$

- \mathbf{Z} : In case of a normally distributed jump, the posterior for the jump size is normally distributed with $\overline{Z_{ih}} \sim \mathcal{N}(B/A, \sqrt{1/A})$ (the two parameters being the mean and the standard deviation) where

$$A = \frac{J_{ih}}{h\sigma^2 X_{(i-1)h}^{2b}} + \frac{1}{\sigma_J^2} \quad \text{and} \quad B = \frac{J_{ih} \left(X_{ih} - (1 - h\kappa)X_{(i-1)h} - h\theta\kappa \right)}{h\sigma^2 X_{(i-1)h}^{2b}} + \frac{\mu_J}{\sigma_J^2}.$$

If jumps are assumed to be exponential, the posterior is a truncated normal $\overline{Z_{ih}} \sim \mathcal{N}(A, B) \mathbb{1}_{\mathbb{R}^+}$ (with support on the positive real axis and the two parameters being the mean and standard deviation of the underlying normal distribution). In the case $J_{ih} = 1$, this distribution is defined by the parameters

$$A = X_{ih} - X_{(i-1)h} + \kappa(\theta - X_{(i-1)h})h + h\sigma^2 \eta_J^{-1} X_{(i-1)h}^{2b} \quad \text{and} \quad B = \sqrt{h}\sigma X_{(i-1)h}^b.$$

If $J_{ih} = 0$, the posterior is a draw from $\overline{Z}_{ih} \sim \exp(\eta_J)$ as the data is uninformative about the jump size (where \exp is the exponential distribution and its parameter is the mean of this distribution).

- κ : We opt for a prior with a normal distribution i.e. $\underline{\kappa} \sim \mathcal{N}(\mu_\kappa, \sigma_\kappa)$. Alternatively, a truncated prior would restrict the support to the positive axis, but since the actual draws for all models are significantly bounded away from zero, a normal prior suffices. This choice implies a normal posterior and thus κ can be simulated directly. The posterior distribution is $\overline{\kappa} \sim \mathcal{N}(B/A, \sqrt{1/A})$, where

$$A = \frac{1}{\sigma_\kappa^2} + \sum_{i=1}^{T/h} \frac{h \left(\theta - X_{(i-1)h} \right)^2}{\sigma^2 X_{(i-1)h}^{2b}} \quad B = \frac{\mu_\kappa}{\sigma_\kappa^2} + \sum_{i=1}^{T/h} \frac{\left(\theta - X_{(i-1)h} \right) \left(X_{ih} - X_{(i-1)h} - J_{ih} Z_{ih} \right)}{\sigma^2 X_{(i-1)h}^{2b}}$$

As it is reasonable to assume that different models imply similar mean-reversion levels, we choose $\mu_\kappa = 0$ and $\sigma_\kappa = 1$, independent of the model. These hyper-parameters imply non-informative prior distributions for all models under consideration.

- θ : We choose a normally distributed prior i.e. $\underline{\theta} \sim \mathcal{N}(\mu_\theta, \sigma_\theta)$. This choice implies a normally distributed posterior and thus it can also be simulated directly. The posterior distribution is $\overline{\theta} \sim \mathcal{N}(B/A, \sqrt{1/A})$, where

$$A = \frac{1}{\sigma_\theta^2} + \sum_{i=1}^{T/h} \frac{h \kappa^2}{\sigma^2 X_{(i-1)h}^{2b}} \quad B = \frac{\mu_\theta}{\sigma_\theta^2} + \sum_{i=1}^{T/h} \frac{\kappa \left(X_{ih} - J_{ih} Z_{ih} + X_{(i-1)h} (h\kappa - 1) \right)}{\sigma^2 X_{(i-1)h}^{2b}}$$

We use similar priors for the log and the level specification. For the level specification we employ $\underline{\theta} \sim \mathcal{N}(18, 5)$, and $\underline{\theta} \sim \mathcal{N}(\log(18), 0.4)$.

- σ : We draw σ^2 and use an inverse Gamma prior, i.e. $\underline{\sigma^2} \sim \mathcal{IG}(\alpha_{\sigma^2}, \beta_{\sigma^2})$. This choice implies an inverse gamma posterior

$$\overline{\sigma^2} \sim \mathcal{IG} \left(\alpha_{\sigma^2} + \frac{T}{2h}, \beta_{\sigma^2} + \frac{1}{2} \sum_{i=1}^{T/h} \frac{\left(X_{ih} - X_{(i-1)h} - \kappa(\theta - X_{(i-1)h})h - Z_{ih} J_{ih} \right)^2}{h X_{(i-1)h}^{2b}} \right).$$

For all models we use distributions with very large variances and hence impose little information. In particular α_{σ^2} and β_{σ^2} are chosen such that their mean is between 0.05^2 and 0.2^2 (depending on the model) and we use a unit variance for all models.

- λ_0 and λ_1 : In the state independent case ($\lambda_1 = 0$), λ_0 can be drawn directly. Since we draw the jumps per discretization interval we draw $\widetilde{\lambda}_0 = h\lambda_0$. Employing a beta distributed prior, i.e. $\underline{\widetilde{\lambda}_0} \sim \mathcal{B}(\alpha_{\widetilde{\lambda}_0}, \beta_{\widetilde{\lambda}_0})$, the posterior is beta as well with

$$\overline{\widetilde{\lambda}_0} \sim \mathcal{B} \left(\alpha_{\widetilde{\lambda}_0} + \sum_{i=1}^{T/h} J_{ih}, \beta_{\widetilde{\lambda}_0} + \frac{T}{h} - \sum_{i=1}^{T/h} J_{ih} \right).$$

where we choose the parameters $\alpha_{\widetilde{\lambda}_0} = 1$ and $\beta_{\widetilde{\lambda}_0} = 15$. In the state-dependent case we

draw the parameters from

$$\overline{(\lambda_0, \lambda_1)} \propto p(\lambda_0, \lambda_1) \sum_{i=1}^{T/h} \left(h \left(\lambda_0 + \lambda_1 X_{(i-1)h} \right) \right)^{J_{ih}} + \left(1 - h \left(\lambda_0 + \lambda_1 X_{(i-1)h} \right) \right)^{1-J_{ih}}$$

using a random walk Metropolis algorithm. The prior $p(\lambda_0, \lambda_1)$ is chosen to be a uniform distribution over $[0, 1] \times [0, 1]$. The right bound is arbitrary but large. This prior provides no information over a wide range of realistic parameter values.

- μ_J and σ_J : These two parameters can be estimated from the time series of estimated jumps \mathbf{Z} . Using a normally distributed prior $\underline{\mu}_J \sim \mathcal{N}(\mu_{\mu_J}, \sigma_{\mu_J})$, the corresponding posterior for the expected jump size is

$$\underline{\mu}_J \sim \mathcal{N} \left(\frac{\sigma_J^2 \mu_{\mu_J} + \sigma_{\mu_J}^2 \sum_{i=1}^{T/h} Z_{ih}}{\sigma_J^2 + \frac{T}{h} \sigma_{\mu_J}^2}, \left(\frac{T}{h \sigma_J^2} + \frac{1}{\sigma_{\mu_J}^2} \right)^{-0.5} \right).$$

Similarly, the jump-size variance can be drawn as a linear regression parameter. Using $\underline{\sigma}_J^2 \sim \mathcal{IG}(\alpha_{\sigma_J^2}, \beta_{\sigma_J^2})$. This choice implies an inverse gamma posterior

$$\underline{\sigma}_J^2 \sim \mathcal{IG} \left(\alpha_{\sigma_J^2} + \frac{T}{2h}, \beta_{\sigma_J^2} + \frac{1}{2} \sum_{i=1}^{T/h} (Z_{ih} - \mu_J)^2 \right).$$

Prior distributions with very large standard deviations are chosen for both jump parameters. We use $\underline{\mu}_J \sim \mathcal{N}(0, 10)$ for the level models and $\underline{\mu}_J \sim \mathcal{N}(0, 1)$ for the log models. For σ_J^2 we utilize a prior distribution with mean 0.01 and standard deviation of 5 for the log models and mean 9 and standard deviation 20 for the level models.

- η_J : We draw $1/\eta_J$. In the case where jumps are exponentially distributed, using a gamma distributed prior, i.e. $1/\eta_J \sim \mathcal{G}(\alpha_{\eta_J}, \beta_{\eta_J})$. The corresponding posterior is gamma as well with

$$\overline{1/\eta_J} \sim \mathcal{G} \left(\alpha_{\eta_J} + \frac{T}{h}, \beta_{\eta_J} + \sum_{i=1}^{T/h} Z_{ih} \right).$$

We use again hyperparameters that imply relatively uninformative jump sizes. In the log models we use parameters that imply a mean of 10 and a standard deviation of 20, whereas in the level models we use a mean of 0.2 and a standard deviation of 1.

B Complete Conditionals and Prior Information - SVV Models

The time discretization between two observations for this model is given by

$$\begin{aligned} X_{t+1} &= X_t + \sum_{i=0}^{1/h-1} \left[\kappa (\theta - X_{t+ih}) h + \sqrt{hV_{t+ih}} \varepsilon_{t+(i+1)h}^x + Z_{t+(i+1)h} J_{t+(i+1)h} \right] \\ V_{t+1} &= V_t + \sum_{i=0}^{1/h-1} \left[\kappa_v (\theta_v - V_{t+ih}) h + \sigma_v \sqrt{hV_{t+ih}} \varepsilon_{t+(i+1)h}^v \right] \end{aligned}$$

where ε_t^x and ε_t^v are both standard normal with correlation ρ . To augment the data set, the draw from the posterior of elements in \mathbf{X}^u is now proportional to

$$p\left(X_{ih} | X_{(i-1)h}, \mathbf{Z}, \mathbf{J}, \mathbf{V}, \Theta\right) \times p\left(X_{(i+1)h} | X_{ih}, \mathbf{Z}, \mathbf{J}, \mathbf{V}, \Theta\right)$$

where we use the same Metropolis algorithm as before.

Prior distributions and the hyper-parameters used in the empirical implementation of this model are given as follows:

- **J**: The posterior distribution of J_{ih} is Bernoulli with jump probability of $p = A/(A+B)$, where

$$\begin{aligned} A &= h\lambda_0 \exp\left[-\frac{1}{2(1-\rho^2)}\left(C_{ih}^2 + (D_{ih}^1)^2 - 2\rho C_{ih}D_{ih}^1\right)\right], \\ B &= (1-h\lambda_0) \exp\left[-\frac{1}{2(1-\rho^2)}\left(C_{ih}^2 + (D_{ih}^0)^2 - 2\rho C_{ih}D_{ih}^0\right)\right]. \end{aligned}$$

where

$$\begin{aligned} C_{ih} &= \left(V_{ih} - V_{(i-1)h} - \kappa_v(\theta_v - V_{(i-1)h})h\right) / \left(\sigma_v \sqrt{hV_{(i-1)h}}\right), \\ D_{ih}^1 &= \left(X_{ih} - X_{(i-1)h} - \kappa(\theta - X_{(i-1)h})h - Z_{ih}\right) / \left(\sqrt{hV_{(i-1)h}}\right), \\ D_{ih}^0 &= \left(X_{ih} - X_{(i-1)h} - \kappa(\theta - X_{(i-1)h})h\right) / \left(\sqrt{hV_{(i-1)h}}\right) \end{aligned}$$

- **Z**: In case of a normally distributed jump, the posterior for the jump size is normally distributed with $\bar{Z}_{ih} \sim \mathcal{N}(B/A, \sqrt{1/A})$ where

$$A = \frac{J_{ih}}{h(1-\rho^2)V_{(i-1)h}} + \frac{1}{\sigma_J^2} \quad \text{and} \quad B = \frac{J_{ih}\left(F_{ih} - \sqrt{h}\rho C_{ih}\sqrt{V_{(i-1)h}}\right)}{h(1-\rho^2)V_{(i-1)h}} + \frac{\mu_J}{\sigma_J^2},$$

where C_{ih} is defined as before and $F_{ih} = X_{ih} - X_{(i-1)h} - \kappa(\theta - X_{(i-1)h})$.

- $\lambda_0, \mu_J, \sigma_J$: These can be estimated based on \mathbf{Z} and \mathbf{J} as before. We use $\tilde{\lambda}_0 \sim \mathcal{B}(1, 25)$ because it is reasonable to assume a priori that the inclusion of a stochastic volatility-of-volatility component reduces the jump intensity. We use priors with little information for the other parameters: $\mu_J \sim \mathcal{N}(0.1, 0.2)$ and for σ_J^2 we use an inverse gamma distribution with mean 0.1^2 and unit standard deviation.
- κ : We use the same priors for this parameter as in the one-dimensional case i.e. $\underline{\kappa} \sim \mathcal{N}(0, 1)$. The posterior for this parameter now reflects the correlation between the two state variables and is given by $\bar{\kappa} \sim \mathcal{N}(B/A, \sqrt{1/A})$, where

$$\begin{aligned} A &= \frac{1}{\sigma_\kappa^2} + \sum_{i=1}^{T/h} \frac{h(\theta - X_{(i-1)h})^2}{(1-\rho^2)V_{(i-1)h}} \\ B &= \frac{\mu_\kappa}{\sigma_\kappa^2} + \sum_{i=1}^{T/h} \frac{(\theta - X_{(i-1)h})\left(-\rho C_{ih}\sqrt{hV_{(i-1)h}} + X_{ih} - X_{(i-1)h} - J_{ih}Z_{ih}\right)}{(1-\rho^2)V_{(i-1)h}}. \end{aligned}$$

- θ : Again using the same prior as in the one dimensional model, the posterior is given by $\theta \sim \mathcal{N}(B/A, \sqrt{1/A})$, where

$$A = \frac{1}{\sigma_\theta^2} + \sum_{i=1}^{T/h} \frac{h\kappa^2}{(1-\varrho^2)V_{(i-1)h}}$$

$$B = \frac{\mu_\theta}{\sigma_\theta^2} + \sum_{i=1}^{T/h} \frac{\kappa \left(-\varrho C_{ih} \sqrt{hV_{(i-1)h}} + (h\kappa - 1)X_{(i-1)h} + X_{ih} - J_{ih}Z_{ih} \right)}{(1-\varrho^2)V_{(i-1)h}}.$$

- κ_v and θ_v : For all models we use $\kappa_v \sim \mathcal{N}(0, 1)$. This leads again to $\bar{\kappa}_v \sim \mathcal{N}(B/A, \sqrt{1/A})$ with

$$A = \frac{1}{\sigma_{\kappa_v}^2} + \sum_{i=1}^{T/h} \frac{h \left(V_{(i-1)h} - \theta_v \right)^2}{(1-\varrho^2)\sigma_v^2 V_{(i-1)h}}$$

$$B = \frac{\mu_{\kappa_v}}{\sigma_{\kappa_v}^2} + \sum_{i=1}^{T/h} \frac{\left(V_{(i-1)h} - \theta_v \right) \left(V_{(i-1)h} - V_{ih} + \varrho D_{ih} \sqrt{hV_{(i-1)h}\sigma_v} \right)}{(1-\varrho^2)\sigma_v^2 V_{(i-1)h}}.$$

Also we have $\bar{\theta}_v \sim \mathcal{N}(B/A, \sqrt{1/A})$ with

$$A = \frac{1}{\sigma_{\theta_v}^2} + \sum_{i=1}^{T/h} \frac{h\kappa_v^2}{(1-\varrho^2)V_{(i-1)h}\sigma_v^2} \quad B = \frac{\mu_{\theta_v}}{\sigma_{\theta_v}^2} - \sum_{i=1}^{T/h} \frac{\kappa_v \left(1 - h\kappa_v + \frac{-V_{ih} + \varrho D_{ih} \sqrt{hV_{(i-1)h}\sigma_v}}{V_{(i-1)h}} \right)}{(1-\varrho^2)\sigma_v^2}.$$

where $D_{ih} = \left(X_{ih} - X_{(i-1)h} - \kappa \left(\theta - X_{(i-1)h} \right) h - Z_{ih}J_{ih} \right) / \left(\sqrt{hV_{(i-1)h}} \right)$. We use $\theta_v \sim \mathcal{N}(0.3, 1)$.

- σ_v and ϱ : Both parameters don't have posterior distributions of known form and hence we update them with a random walk Metropolis algorithm. We assume uniform priors for both parameters, where the parameter ϱ is restricted to $[-1, 1]$ and σ_v is restricted to $[0, 1]$. For σ_v one could alternatively choose a prior with support on the whole real axis.
- \mathbf{V} : We update the variance one at a time. This implies the following full conditional distribution for V_{ih} :

$$p(V_{ih} | \mathbf{V}_{-ih}, \mathbf{X}, \mathbf{Z}, \mathbf{J}, \Theta) \propto \frac{1}{V_{ih}} \exp \left[-\frac{1}{2(1-\varrho^2)} \left(C_{ih}^2 + D_{ih}^2 - 2\varrho C_{ih}D_{ih} \right) \right]$$

$$\times \exp \left[-\frac{1}{2(1-\varrho^2)} \left(C_{(i+1)h}^2 + D_{(i+1)h}^2 - 2\varrho C_{(i+1)h}D_{(i+1)h} \right) \right]$$

where C_{ih} and D_{ih} are defined as before and \mathbf{V}_{-ih} denotes the variance vector except for V_{ih} . For the first and the last day in the sample the formulae apply with a slight adjustment. We use a random walk Metropolis algorithm, tuned to yield acceptance rates of around 40%.

References

- Bayarri, M. and Berger, J. (2000). P-values for composite null models. *Journal of the American Statistical Association*, **95**, 1127–1142.
- Becker, R., Clements, A. E., and McClelland, A. (2009). The jump component of S&P 500 volatility and the VIX index. *Journal of Banking and Finance*, **33**(6), 1033–1038.
- Britten-Jones, M. and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *Journal of Finance*, **55**(2), 839–866.
- 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.
- 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., 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*, forthcoming.
- Detemple, J. B. and Osakwe, C. (2000). The valuation of volatility options. *European Finance Review*, **4**(1), 21–50.
- Dotsis, G., Psychoyios, D., and Skiadopoulos, G. (2007). An empirical comparison of continuous-time models of implied volatility indices. *Journal of Banking and Finance*, **31**(12), 3584–3603.
- Duan, J.-C. and Yeh, C.-Y. (2010). Jump and volatility risk premiums implied by VIX. *Journal of Economic Dynamics and Control*, **In Press, Corrected Proof**, –.
- Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, **19**(2), 177–91.
- Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, **53**(3), 1269–1300.
- Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**, 733–807.
- Geman, S. and Geman, G. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*. John Wiley and Son.
- Grunbichler, A. and Longstaff, F. A. (1996). Valuing futures and options on volatility. *Journal of Banking & Finance*, **20**(6), 985–1001.
- Hagan, P., Kumar, D., Lesniewski, A., and Woodward, D. (2002). Managing smile risk. *Wilmott Magazine*, **1**, 84–108.

- Hull, J. C. and White, A. D. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, **42**(2), 281–300.
- 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. (2005). The model-free implied volatility and its information content. *Review of Financial Studies*, **18**(4), 1305–1342.
- 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. (2004). The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance*, **59**(1), 227–260.
- Jones, C. (1998). Bayesian estimation of continuous-time finance models. *Unpublished Working Paper, University of Rochester*.
- 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. (2010). Stochastic volatility jump-diffusions for equity index dynamics. *ICMA Centre Discussion Papers, DP 2010-06*.
- Konstantinidi, E., Skiadopoulos, G., and Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking and Finance*, **32**(11), 2401–2411.
- Meng, X. L. (1994). Posterior predictive p-values. *Ann. Statist.*, **22**, 1142–1160.
- Psychoyios, D., Dotsis, G., and Markellos, R. N. (2010). A jump diffusion model for VIX volatility options and futures. *Review of Quantitative Finance and Accounting, forthcoming*.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, **12**, 1151–1172.
- Whaley, R. (1993). Derivatives on market volatility: Hedging tools long overdue. *Journal of Derivatives*, **1**, 71–84.
- Wu, L. (2010). Variance dynamics: Joint evidence from options and high-frequency returns. *Journal of Econometrics*, **doi:10.1016/j.jeconom.2010.03.037**.