Inferring Volatility Dynamics and Risk Premia from the S&P 500 and VIX Markets

Chris Bardgett          Elise Gourier
Markus Leippold

First version: May 2012
Current version: July 2013

This research has been carried out within the NCCR FINRISK project on “New Methods in Theoretical and Empirical Asset Pricing” and “Credit Risk and Non-Standard Sources of Risk in Finance”
Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets∗

CHRIS BARDGETT† ELISE GOURIER‡ MARKUS LEIPPOLD§

July 21, 2013

Abstract

This paper provides a thorough analysis of the empirical performance of affine jump-diffusion models to jointly represent the values of the S&P 500 and VIX indices, and options on both markets with wide ranges of maturities and moneynesses. Based on the affine relationship of the VIX squared with respect to the latent factors, we extend the Fourier Cosine Expansion to efficiently price VIX derivatives. Using efficient filtering methods, we investigate the behavior of the latent processes as well as the out-of-sample performance of sub-models depending on which products and markets are considered in the in-sample estimation procedure. We find that a stochastic central tendency improves significantly the representation of the tails of the returns’ distribution and the term structure of the smiles of volatility in both the S&P 500 and VIX markets. Furthermore, jumps in volatility help reproduce the change of regime in the VIX derivatives market during the crisis and the tail of the variance risk-neutral distribution. We analyze the information contents of the S&P 500 and VIX markets and argue that they provide complementary information to estimate jumps, therefore calibrating models to only one market is not sufficient to price options on the other market. Finally, we investigate the risk premia of each factor and find that the positive equity risk premium is mainly determined by its continuous part, whereas the variance risk premium is only slightly negative and mainly affected by its jump part.

Keywords: S&P 500 and VIX joint modeling, option pricing, particle filter, dynamics of volatility.

∗The authors thank Fabio Trojani, Jérôme Detemple, Josef Teichmann, Chris Rogers, Anders Trolle, Alexandre Ziegler, Garland Durham, Andras Fulop, Peter Christoffersen and the seminar participants of the Brown Bag Lunch Seminar at the Department of Banking and Finance Institute at the University of Zürich, the Gerensce Swiss Finance Institute PhD workshop, the Bachelier 2012 Finance Conference, the joint seminar with ETH Zürich and the 5th International Conference of the ERCIM WG on Computing & Statistics for helpful comments. Financial support from the Swiss Finance Institute (SFI), Bank Vontobel, and the National Center of Competence in Research “Financial Valuation and Risk Management” is gratefully acknowledged.

†University of Zürich and Swiss Finance Institute (SFI), Plattenstrasse 14, 8032 Zürich, Switzerland; tel: (+41)-44-634-4045; Email: chris.bardgett@bf.uzh.ch.

‡University of Zürich and Swiss Finance Institute (SFI), Plattenstrasse 14, 8032 Zürich, Switzerland; tel: (+41)-44-634-4045; Email: elise.gourier@bf.uzh.ch.

§University of Zürich and Swiss Finance Institute (SFI), Plattenstrasse 14, 8032 Zürich, Switzerland; tel: (+41)-44-634-5069; Email: markus.leippold@bf.uzh.ch.
1 Introduction

One of the central questions addressed by research in empirical option pricing is the determination of asset returns dynamics. Ideally, in addition to reproducing asset prices and prices of derivatives traded on a given day on the market, they should also recreate the joint evolution of these prices over time. Under the historical measure $\mathbb{P}$ the time series of asset returns provide valuable information on the main characteristics of returns dynamics (assuming stationarity and ergodicity). On the other hand option prices on this asset help to specify its dynamics under the risk neutral measure $\mathbb{Q}$. Indeed, the result of Breeden and Litzenberger (1978) states that the observation of vanilla option prices with maturity $T$ for a continuum of strikes entirely determines the $\mathbb{Q}$ distribution of this asset at the future time $T$. Even though we do not observe prices for arbitrary strikes in practice, the S&P 500 index has many strikes traded liquidly. Whereas a large part of the literature on asset pricing either focuses on the time-series properties of returns under the historical measure or proposes models to accurately capture the stylized facts of option prices, the study of the link between both measures has recently captured more attention. The change of measure from $\mathbb{P}$ to $\mathbb{Q}$ is achieved through an appropriate specification of risk premia, which can be interpreted as compensations for the risks that investors take when buying an asset. While there is a large amount of research articles and surveys on the equity risk premium, the study of the variance dynamics and variance risk premium is more recent. Prominent examples include Bates (1996, 2000, 2003), Chernov and Ghysels (2000), Jackwerth (2000), Pan (2002), Jones (2003), Eraker (2004), Aït-Sahalia and Kimmel (2007), Broadie, Chernov, and Johannes (2007), Carr and Wu (2009), Todorov (2010) and Wu (2011). However, the components of risk premia, in particular when jumps are involved, are usually found hard to estimate and statistically insignificant with daily data. One reason for this is that the estimation of risk premia requires a large amount of returns and options data and therefore powerful computational tools to extract the relevant information. In fact, because of this computational burden, most research does not make use of the whole cross section of options and considerably reduces the amount of information. As available sources of information have grown tremendously since the introduction of the volatility index VIX as well as VIX derivatives, the need for efficient computational algorithms arises.

The VIX index has been constructed to approximate non-parametrically the expected future realized volatility of the S&P 500 returns over the next 30 days. The VIX is not directly tradeable but it
is possible to trade VIX futures and options. The options started trading in 2006 and have been a growing business at a dramatic speed ever since. They now represent a much larger market than VIX futures. By definition, the VIX index is linked to the dynamics of the S&P 500 index returns and this makes VIX and S&P 500 options both ideal to infer these dynamics. Fortunately, the S&P 500 and VIX options markets are among the most liquid worldwide with a daily average volume of 783,768 and 391,992 contracts traded per day in 2011, and therefore represent a trustworthy source of information. Including more information on volatility and its evolution over time is essential to better specify and understand the dynamics of volatility. Subsequently, we will use interchangeably S&P 500 and SPX, which is its ticker symbol.

Our contribution is the following: We develop an algorithm that uses time series of returns and derivatives on both the SPX and VIX markets to investigate the historical and risk-neutral dynamics of the SPX returns. First of all, we base our analysis on a time series of cross sectional data of options with a wide range of moneynesses and maturities. We emphasize that we have kept liquid deep out-of-the-money options in our dataset in contrast to most of the literature, so as to keep valuable information about the tails of S&P 500 returns. These options are considered to be essential in the estimation of the jump structure of their underlying under the risk-neutral measure, i.e., of the jumps of the S&P 500 and VIX indices. Indeed, the steepness of the S&P 500 smile and the high volatilities for short maturities puts is considered to be a strong indication of jumps in the returns. Similarly, the positive skewness of VIX implied volatilities and high volatilities of deep out-of-the-money calls can be explained by the presence of positive jumps in the VIX. Therefore, the cross section of options is required to infer possible jumps under the risk-neutral measure and justifies why we have decided to use the whole cross section of S&P 500 and VIX options. Second, including derivatives on both the S&P 500 and the VIX indices allows us to make better inference on model parameters and risk premia dynamics, which are consistent with both markets. Third, we analyze the information contained in the S&P 500 and VIX derivatives markets separately, and then examine how models perform to incorporate information from both markets together. In particular, we discuss how the dynamics of variance and its central tendency can be different when estimating models to different derivatives markets. We also study how models estimated to S&P 500 options perform in pricing VIX derivatives and vice-versa. This provides an important analysis of the information content of the S&P 500 and VIX options markets. Up to now and to our knowledge, extracting information from both SPX and VIX derivatives markets has not been done and therefore provides new valuable
insight into the dynamics of asset returns and volatility.

We model the S&P 500 returns using the affine framework of Duffie, Pan, and Singleton (2000). This structure allows us to price S&P 500 and VIX derivatives in semi-closed form and is essential to carry out the analysis of returns and volatility dynamics using such a large dataset of options. However, we point out and reduce the limitations of one-factor affine models by advocating a stochastic level of reversion in the volatility dynamics, which is a key ingredient in order to consistently accommodate for the time series of both markets. The flexibility of this model helps to investigate how many factors are needed to reproduce the times series features of the data, and whether jumps should be incorporated.

Estimating the dynamics using such an extremely large dataset of options on the two markets and for a long time series requires computationally efficient techniques that can easily deal with the complicated features of the model, in particular state-dependent jumps. To achieve this goal, we extend the Cosine method introduced by Fang and Oosterlee (2008) for S&P 500 options to price VIX options and adapt the Auxiliary Particle Filter of Pitt and Shephard (1999) to filter out unobservable processes over time and their jumps. Sequential Monte-Carlo techniques have been recently used for limited datasets, but most papers restrict their dataset of options to near at-the-money options. Furthermore, we want to stress that our estimation methodology consists in a single step using the times series of indices and options together. In particular, we do not estimate the model under the historical measure and then fix parameters to estimate the risk-neutral dynamics. Our approach increases the computational complexity but ensures a consistent estimation of the historical and pricing measures, which is essential to estimate reliably risk premia.

Among the outputs of the Auxiliary Particle Filter are the filtered trajectories of the latent processes. We investigate their behavior, compare them across models and when estimating to different datasets. This analysis allows us to draw conclusions regarding the usefulness of jumps and the number of factors needed to represent volatility. Furthermore, we investigate the information content of the underlying levels and of the options on each market, compare them and provide a discussion on risk premia.

The paper is organized as follows. In Section 2, we explain how our paper fits into the existing literature. We then conduct a preliminary data analysis in Section 3, and highlight some differences
between the S&P 500 and the VIX options markets. In Section 4 we present the two-factor affine jump-diffusion model that we use later in the estimation. We describe the risk premium specification and derive the pricing formula for the VIX squared as well as VIX and S&P 500 options. In Section 5 we discuss the joint estimation to one single day of data as well as the Auxiliary Particle Filter that we use to calibrate the model to a time-series of cross-sectional options data. Finally, in Sections 6 and 7 we summarize the results of the daily and time-series estimations and present our findings. Section 8 concludes.

2 Related literature

Our work builds on an extensive body of research that analyzes which features are needed for a model to provide a realistic representation of equity underlying and derivatives prices. While the end of the twentieth century has been characterized by a fast growing literature on equity option pricing, the financial crisis has recently drawn more attention to the need to better understand and model equity volatility. Since the introduction of the volatility index VIX in 1993 and its derivatives (from 2004 onwards), the direct modeling of volatility and the pricing of its derivatives has been the focus of numerous papers. We refer among others to Whaley (1993), Grünbichler and Longstaff (1996), Detemple and Osakwe (2000), Bergomi (2004, 2005, 2008), Sepp (2008a,b), Bergomi (2009), Lian and Zhu (2011), Drimus and Farkas (2013) and Mencía and Sentana (2013). An important conclusion of this literature is that sharp increases in the variance dynamics are necessary to reproduce the positive skewness of VIX options’ implied volatilities. In particular, many articles point out that this could be achieved by having positive jumps in the variance. In particular, Christoffersen, Jacobs, and Mimouni (2010) demonstrate via Q-Q plots that using a square-root model without jumps for the variance is not in line with empirical properties of the data. Using realized variance taken at high-frequency as a proxy for the integrated variance, they show that the empirical realized volatility is not Gaussian as the continuous square-root model posits. Jumps allow the distribution of the integrated volatility to be fatter-tailed and therefore represent better the data. Todorov (2010) and Todorov and Tauchen (2011) test for jumps in the VIX index and finds strong evidence supporting this assumption. He also tests for co-jumps in S&P 500 returns and in the VIX and finds striking evidence for them. He finally finds that 63% of the co-jump variation in the sample studied is due to the combination of negative jumps in the returns and positive jumps in the volatility. Jacod and
Todorov (2010) develop further statistical tests which indicate that most stock market jumps are associated with volatility jumps. Eraker, Johannes, and Polson (2003) show that using jumps in the volatility process significantly improves the fit of returns. Finally, as mentioned in Eraker (2004), continuous volatility or variance processes are not able to explain the unusually large volatility before and after the crash of 1987. The specification of jumps is furthermore of importance. Bates (1996), Pan (2002) and Eraker (2004) argue in favor of using state-dependent jumps in returns, which is intuitively appealing as jumps tend to occur more frequently when volatility increases. Using variance swaps, Aït-Sahalia, Karaman, and Mancini (2012) found that the state dependent intensity of jumps was a desirable model feature. However, evidence supporting this choice is mixed. Indeed, Bates (2000) finds that state dependent intensities lead to strong misspecification and Eraker (2004) finds that it does not significantly improve the option prices fit. Broadie, Chernov, and Johannes (2007) and Johannes, Polson, and Stroud (2009) use a constant intensity of jumps.

Another concern of volatility modeling relates to the number of factors that should be used. While adding an additional factor to the Heston model increases the complexity, it has indeed been shown that two factors are needed to provide an accurate description of the volatility dynamics (see, e.g., Andersen, Benzoni, and Lund (2002), Alizadeh, Brandt, and Diebold (2002), Chernov, Gallant, Ghysels, and Tauchen (2003), Todorov (2010), Kaeck and Alexander (2012), Bates (2012) and Mencía and Sentana (2013)).

Several papers have been published in the last years aiming to reconcile the cross-sectional information of the S&P 500 and the VIX derivatives markets by modeling them jointly. Gatheral (2008) pointed out first that even though the Heston model performs fairly well to price S&P 500 options, it totally fails to price VIX options. Figure 4 shows that modeling the instantaneous volatility as a square root process leads to a VIX smile decreasing with moneyness, which is the opposite of what is observed in practice. Therefore the volatility density implied by VIX options has more mass at high volatility and less mass at lower volatility levels than the Chi-Square density of the Heston model. Some studies are going in the direction of non-affine models (e.g., Jones (2003), Aït-Sahalia and Kimmel (2007), Christoffersen, Jacobs, and Mimouni (2010), Ferriani and Pastorello (2012), Durham (2012), Kaeck and Alexander (2012)). However tractability remains an issue that is of crucial importance when it comes to calibrating a model to a long time series containing hundreds of options each day.
Among the recent papers that attempted to reproduce simultaneously the smiles of volatility of S&P 500 and VIX options are Chung, Tsai, Wang, and Wenig (2011), Cont and Kokholm (2011), Song and Xiu (2012), Papanicolaou and Sircar (2012) and Bayer, Gatheral, and Karlsmark (2013). We build on this literature by considering extensions of the Heston model that remain in the affine framework, but add more flexibility to the specifications used in the above mentioned papers. Our model is a special case of the general affine framework developed by Duffie, Pan, and Singleton (2000) but includes as sub-cases the usual extensions of the Heston model encountered in the literature, for example Bates (2000), Eraker (2004) and Sepp (2008a).

However most if not all of the papers that consider S&P 500 and VIX options in their calibration exercise have restricted their analysis to a static one-day estimation. Therefore the estimated parameters might exhibit large variations when calibrating the model to different dates. Lindström, Ströjby, Brodén, Wiktorsson, and Holst (2008) show that the estimated parameters are not stable over time and therefore cannot be used to infer time series properties of returns and risk premia. In the last decade, powerful algorithms have been developed to estimate non-linear models with non-Gaussian innovations in a time-consistent manner.

Time-consistent estimation methods have been used so far to calibrate models to index returns and options. For example, Pan (2002) uses a tailored version of the Generalized Methods of Moments to estimate the Bates model using a time series of S&P 500 and options (two per day). Eraker (2004) relies on Markov Chain Monte Carlo methods to estimate risk premia for jumps in returns and volatility also using returns and options (around three per day). Broadie, Chernov, and Johannes (2007) were the first to consider the whole cross section of option prices on the S&P 500. To reduce the computational burden, they fix some of the parameters by taking values from previous estimations of the time series of returns and minimize a least square type distance between market and model option implied volatilities. They find that the time series provided evidence that volatility jumps, which coincides with the literature that appeared later on VIX option pricing. With a particle filter, Johannes, Polson, and Stroud (2009) investigate whether the time-series of returns of the S&P 500 are consistent with information embedded in option prices. Their options sample is limited to one option per day. They find some inconsistencies that they attribute to either a wrong specification of risk premia or a lack of flexibility of the model. They conclude that their results might be explained by the introduction of a time varying level of reversion for the volatility. Christoffersen, Jacobs,
and Mimouni (2010) apply a Maximum Likelihood Importance Sampling technique on returns and a separate Non-linear Least-Squares Important Sampling estimation to option prices to compare the accuracy of models in reproducing returns and option prices. However, as underlined in Ferriani and Pastorello (2012), most papers filtering information from option prices rely on one option per day or a very limited set of options. This is computationally less intensive but ignores a large part of the information present on the market. Ferriani and Pastorello (2012) have used part of the cross section of options and the time series of log-returns in the filtering problem. They do not consider jumps in the volatility but study different non-affine models. They conclude that significant improvement could be brought into these models by incorporating jumps or regime switching in the volatility dynamics. Finally, in a working paper Duan and Yeh (2011) use a filter on the S&P 500 returns together with the VIX index to infer the dynamics of returns and volatility. However, they do not use options data making it impossible to estimate risk premia.

3 Preliminary data analysis

In Figure 1, we plot the joint evolution of the S&P 500 and the VIX index. Their movements are highly negatively correlated, which explains the use of instruments on the VIX to hedge part of the equity risk of a portfolio. Table 1 displays the first four moments of the S&P 500 returns and VIX index levels, over two periods of time. The first period covers from March 2006 until November 2008, i.e., it spans the pre-crisis period as well as the beginning of the crisis. The second period starts in December 2008 and lasts until October 2010. Log-returns on the S&P 500 exhibit negative skewness during the second period considered, and a high kurtosis over both periods, suggesting the presence of rare and large movements. The VIX index exhibits a large skewness and kurtosis in the first period, but in the second period the statistics suggest that the movements are more symmetric, centered about a higher value (29% instead of 20% in the first period).

[Insert Figure 1 here]

[Insert Table 1 here]

We consider closing prices of European options on the S&P 500 from March 1, 2006 to October 29, 2010. The data was obtained from OptionMetrics. The time period of our dataset is restricted by
the fact that options on the VIX were introduced in 2006. We also use a dataset of VIX options closing prices on the same time period coming from the data provider DeltaNeutral. This time series includes periods of calm and periods of crisis with extreme events, especially relevant to estimate the presence and magnitude of jumps. In particular, during the financial crisis that started at the beginning of 2007, the VIX index was at its highest peak since its launch.

Both the S&P 500 and VIX options dataset are treated following usual procedures (see Aït-Sahalia and Lo (1998)). In particular, we only consider options with maturity between one week and one year and delete options quotes that were not traded on a given date. We follow two main steps. First, we delete all in-the-money (ITM) options since they are illiquid compared to out-of-the-money (OTM) options. Second, we infer from highly liquid options the Futures price using the at-the-money (ATM) put-call parity. This avoids two issues: Making predictions on future dividends, and using Futures closing prices which are not synchronized with the option closing prices. Hence, we consider that the underlying of the options is the index Futures and not the index itself. At the end, we only work with liquid OTM options for the S&P 500 market and only with liquid call options for the VIX market. Indeed, in the case where the VIX ITM call is not liquid, we use the put-call parity to infer from a more liquid VIX OTM put a liquid VIX ITM call.

These adjustments leave a total of 383,286 OTM S&P 500 options and a total of 43,775 call options on the VIX. This implies a daily average of 327 S&P 500 options and 37 VIX options. The number of S&P 500 options in our dataset on a given date is increasing with time with around 170 options at the beginning of the dataset and around 450 options at the end. For VIX options, the number is increasing substantially, with around 5 options per day at the beginning and around 70 options per day at the end. At the beginning of the sample, there are one or two short maturities (below 6 months) available for VIX options and around 6 maturities for S&P 500 options with approximately 40 options per maturity slice. At the end of the sample, VIX options have around 5 short maturities (less than 6 months) with a bit more than 10 options trading per maturity. For S&P 500 options, around ten maturities are available per day with around 60 options for one-month maturities and 40 options for the one-year slice. The low number of VIX options compared to the number of S&P 500 options is first coming from the fact that the VIX options market started in 2006 and therefore that the overall volume traded is lower but also from the fact that less maturities and less strikes are traded. At the end of our sample, the total VIX options volume per day is about half the total.
volume of S&P 500 options traded but much fewer strikes are traded for VIX options.

It is important to understand that calculating implied volatilities of VIX options using as underlying the VIX index is incorrect. Indeed, the true underlying of VIX options is the VIX Futures value. This can intuitively be explained by the fact that a call option at time $t$ with maturity $T$ is an option on volatility on the time interval $[T, T + 30d]$, where $30d$ stands for 30 days. The value $\text{VIX}_t$ at time $t$ is related to volatility on the time interval $[t, t + 30d]$ which might not overlap at all with $[T, T + 30d]$. On the contrary, a Futures on the VIX with maturity $T$ is based on the volatility on the time interval $[T, T + 30d]$. This remark is important because traded VIX option prices do not satisfy no-arbitrage relations with respect to VIX index, but rather with respect to the VIX Futures value. In particular, calculating implied volatilities assuming that the underlying is the VIX might lead to volatilities equal to zero, or which simply do not exist. For this reason all implied volatilities are calculated with respect to the Futures price of the VIX. The same is done for S&P 500 options as it eliminates the need to make predictions on futures dividends.

Even though the S&P 500 and VIX markets are related, we want to emphasize that VIX options behave in a completely different way than S&P 500 options. First, S&P 500 and VIX derivatives with the same maturity contain different information. On the one hand, an S&P 500 option with maturity $T$ contains information about the future S&P 500 index level at time $T$ and therefore about the S&P 500 volatility up to $T$. On the other hand, a VIX option with maturity $T$ embeds information about the VIX at time $T$ and therefore about the S&P 500 volatility between $T$ and $T + 30$ days. Second, the implied volatility smiles backed out from S&P 500 and VIX option prices have very different shapes. Figure 2 displays the S&P 500 and VIX smiles depending on different states of the economy. These implied volatilities (IVs) are computed using the Black-Scholes formula, i.e., backing out the standard deviation of a log-normal distribution for the S&P 500 index (respectively for the VIX index) that are implied by their respective option prices. The VIX IVs are in general substantially higher - ranging from 40% to 200%, with an average IV of around 75% (see Table 2) - than S&P 500 IVs (average IV of around 23%). The implied volatilities are negatively skewed for S&P 500 options, generally decreasing with moneyness as risk-averse investors require a premium for negative states of the economy. In contrast, VIX implied volatilities are positively skewed and increase with moneyness, which can intuitively be explained by the fact that negative returns are often observed together with a rise of volatility (the so-called leverage effect) also corresponding to turbulent states.
of the economy.

[Insert Figure 2 here]

[Insert Table 2 here]

The difference between these markets is also reflected by other indicators such as the put-call trading ratio: Almost twice as many puts as calls are traded daily in the S&P 500 options market but the situation is reversed in the VIX market where the amount of calls traded daily is almost the double of that of the puts. In fact, one can additionally see in Figure 2 that the log-moneynesses traded for S&P 500 options are mostly negative (which corresponds to out-of-the-money put options) and often positive for VIX options (out-of-the-money calls).

Figure 3 represents the expected forward returns of the underlying S&P 500 index returns from March 1st, 2006 to October 29th, 2010 as implied by prices of S&P 500 options with maturity 1 month. We use the method described in Bakshi, Kapadia, and Madan (2003) to calculate the moments implied by option prices. The expected forward returns illustrates the variety of market situations that our time series includes. They were almost constant until the end of 2007, equal to a positive value and thus indicating that market participants were expecting a stable income from investing in the index. But from the end of 2007 they exhibit more variation and seem to mean-revert around a negative trend. Suddenly, following the bail out of Lehman Brothers in September 2008, expected forward returns drop and reach -2% beginning of October 2008. Then they gradually come back and stabilize in mid-2009 around a slightly negative level close to -0.2%. In 2010, the sudden increase in the VIX index coincides with a peak of the expected forward returns reaching about -0.8%. It is interesting to compare the VIX index and expected forward returns as implied by S&P 500 options. Indeed, both indicate market expectations over the next month as reflected in index option prices. However volatility provides information on returns through the leverage effect, while the implied expected forward returns are a direct measure of how investors expect returns to behave. They are much more stable in quiet periods and better reflect the different market situations that compose our time-series and that we aim to reproduce with a model.

[Insert Figure 3 here]
Panel B of Figure 3 displays the expected forward returns on the VIX as implied by VIX options. They remain negative throughout the time series, with large peaks that occur simultaneously with the two peaks of the VIX, reaching between -20% and -25%.

4 Model and option pricing

In this section we present the three-factor affine model that we use. This class of models is known to yield semi-closed form expressions for the price of European options on the S&P 500 index. We show that an additional advantage of this model is that the VIX squared can be expressed as an affine function of the variance and of its level of mean reversion. This allows us to use Fourier analysis and derive semi closed-form expressions for the prices of European claims on the VIX as well.

4.1 Model specification

We consider a filtered probability space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})\) satisfying the usual assumptions, where \(\mathbb{P}\) denotes the historical measure. We fix a risk-neutral measure \(\mathbb{Q}\) equivalent to \(\mathbb{P}\) and denote by \((F_t)_{t \geq 0}\) the forward price\(^1\) of the S&P 500 index and by \(Y_t = (\log(F_t))_{t \geq 0}\) the returns. The dynamics of \(Y\) under \(\mathbb{Q}\) are specified as follows:

\[
\begin{align*}
    dY_t &= [-\lambda X v_t (v_t - m_t) (\theta_Z^{(Q)} (1, 0, 0) - 1) - \frac{1}{2} \sqrt{v_t} - dW_t^{Y(Q)} + dJ_t^{Y(Q)}] dt + \sqrt{v_t} - dW_t^{Y(Q)} + dJ_t^{Y(Q)} \\
    dv_t &= \kappa_v^{(Q)} (m_t - v_t) dt + \sigma_v \sqrt{v_t} - dW_t^{v(Q)} + dJ_t^{v(Q)} \\
    dm_t &= \kappa_m^{(Q)} (\theta_m^{(Q)} - m_t) dt + \sigma_m \sqrt{m_t} - dW_t^{m(Q)} + dJ_t^{m(Q)}
\end{align*}
\]

where \(W^Y, W^v, W^m\) are three \(\mathbb{Q}\) Brownian motions and

\[
    d\langle W^Y, W^v \rangle_t = \rho_{Y,v} dt ; \quad d\langle W^m, W^Y \rangle_t = 0 ; \quad d\langle W^m, W^v \rangle_t = 0.
\]

\(^1\)Assuming that the interest rate \(r\) and dividend yield are constant, it does not matter which maturity of the forward we consider because the cash-and-carry relationship between the forward and the spot index ensures that all forwards have the same dynamics (but different initial conditions).
Our model is a two-factor stochastic volatility model with jumps, which allows the variance process $(v_t)_{t \geq 0}$ of the forward returns to revert towards a stochastic central tendency $(m_t)_{t \geq 0}$. Egloff, Leippold, and Wu (2010) show that this model provides an improvement over the one-factor model in pricing variance swaps. The processes $J^Y, J^v, J^m$ are finite activity jump processes defined by:

$$dJ_t^Y = Z_t^{Y(Q)} dN_t^{Y^v}; \quad dJ_t^v = Z_t^{v(Q)} dN_t^{Y^v}; \quad dJ_t^m = Z_t^{m(Q)} dN_t^m. \quad (5)$$

As suggested by the simultaneous peaks in the S&P 500 and VIX index, and in the expected forward returns on both indices, large movements in the equity returns and in the variance are likely to occur at the same time. Therefore we choose, in line with the literature (see, e.g., Eraker (2004), Broadie, Chernov, and Johannes (2007), Cont and Kokholm (2011)) to use the same Poisson process to generate jumps in the asset returns and in the variance process. We also choose the intensity of jumps to be dependent on the level of the factors. Formally, $N_t^m$ and $N_t^{Y^v}$ are Poisson processes with respective intensities:

$$\lambda^m(m_t-) = \lambda_0^m + \lambda_1^m m_t- \quad (6)$$

$$\lambda^{Yv}(v_t-, m_t-) = \lambda_0^{Yv} + \lambda_1^{Yv} v_t- + \lambda_2^{Yv} m_t- \quad (7)$$

Moreover, the 3-dimensional process $Z^{(Q)} = (Z^{Y(Q)}, Z^{v(Q)}, Z^{m(Q)})^\top$ (where $\cdot^\top$ denotes the transpose operator) corresponds to the random jump sizes under $Q$ and we assume that their values taken at two times $t$ and $s$ are independent and identically distributed (i.i.d.) for all $t \neq s$. We assume that jump sizes in the forward returns are normally distributed $\mathcal{N}(\mu^{(Q)}_{Y^v}, \sigma^{(Q)}_{Y^v})$ and that the jump sizes in the two volatility factors are exponentially distributed with respective means $\nu^{(Q)}_v$ and $\nu^{(Q)}_m$. All jump sizes are independent from one another. These jumps sizes are characterized by their joint Laplace transform:

$$\theta^{(Q)}_Z(\phi) = \theta^{(Q)}_Z(\phi_Y, \phi_v, \phi_m) = \mathbb{E}^Q[\exp(\phi^\top Z^{(Q)})], \quad (8)$$

where $\phi \in \mathbb{C}^3$. 

13
This model implicitly defines dynamics for the VIX. The VIX index is formally defined in the white paper of the CBOE (2009) and is calculated in practice using a combination of S&P 500 options with maturities adjacent to 30 days. Intuitively, the VIX squared is close to the 30-day expected future realized variance and therefore the value of the VIX index should be close to the 30 day-variance swap on the S&P 500 returns. Demeterfi, Derman, Kamal, and Zou (1999) showed that variance swaps can be partially hedged (and therefore priced) using a combination of vanilla options and this is where the formal definition of the VIX is coming from.

In the following, we do not make the assumption that the VIX is approximately the 30-day realized volatility. Instead, we use its definition as a finite sum of call and put prices that converges (under the assumption that there exists call and put options for all strikes in \( \mathbb{R}_+ \)) to the integral

\[
VIX^2_t = \frac{2}{\tau} \mathbb{E}_t^Q \left[ \int_t^{t+\tau} \frac{dF_u}{F_u} - d(\ln F_u) \right]
= \frac{1}{\tau} \mathbb{E}_t^Q \left[ \int_t^{t+\tau} v_u dt + 2 \left( e^{Z^Y_{u}(Q)} - 1 - Z^Y_{u}(Q) \right) dN^Y_{v} \right],
\]

where \( \tau \) is 30 days in annual terms.

In the affine model we use, the expression of the VIX can be shown to be:

**Proposition 4.1.** The VIX squared at time \( t \) can be written as an affine function of \( v_t \) and \( m_t \):

\[
VIX^2_t = \alpha_{VIX} v_t + \beta_{VIX} m_t + \gamma_{VIX}^2
\]

where the coefficients \( \alpha_{VIX} \), \( \beta_{VIX} \) and \( \gamma_{VIX}^2 \) are known in closed-form.

The coefficients \( \alpha_{VIX} \), \( \beta_{VIX} \) and \( \gamma_{VIX}^2 \) are provided in the Appendix A.

### 4.2 Risk premium specification

We specify the change of measure from the pricing to the historical measure so that the model dynamics keep the same structure under \( \mathbb{P} \). The parameters under \( \mathbb{P} \) will simply have a superscript referring to the historical measure. Similarly to Broadie, Chernov, and Johannes (2007), we separate the total equity risk premium \( \gamma_t \) into a Brownian contribution which is proportional to the variance level and represents the compensation for the diffusive price risk, and a jump contribution which
reflects the compensation for jump risk:

\[ \gamma_t = \eta v_t - + \lambda \nu v_t - \nu_m - \theta_{Z_{(P)}} (1, 0, 0) - \theta_{Z_{(Q)}} (1, 0, 0). \]  

(10)

where \( \theta_{Z_{(P)}} \) denotes the joint Laplace transform of jump sizes under the historical measure \( \mathbb{P} \).

As in Pan (2002) and Eraker (2004) we impose the intensity of jumps to be the same under \( \mathbb{Q} \) and \( \mathbb{P} \).\(^2\)

We define the mean price jump risk premium as the difference between the mean of the jump sizes in returns under \( \mathbb{Q} \) and \( \mathbb{P} \). Analogously, the volatility of price jump risk premium refers to the difference between the volatility of the jump sizes in returns under \( \mathbb{Q} \) and \( \mathbb{P} \).\(^3\)

We proceed similarly with the volatility risk premium and decompose it into a diffusive component and a jump component. The diffusive variance risk premium in \( v \) is proportional to the current level of variance, with coefficient of proportionality given by:

\[ \eta_v = \kappa_{v_{(Q)}} - \kappa_{v_{(P)}}. \]  

(11)

This risk premium should primarily be identified by the term structure of SPX implied volatilities as well as the cross section of VIX implied volatilities. The jump part of the volatility risk premium refers to the difference between the mean of the jump sizes in the variance under \( \mathbb{Q} \) and \( \mathbb{P} \).

Finally, we introduce a risk premium in the stochastic central tendency, which consists of a diffusive part proportional to the variance of \( m \) with coefficient of proportionality given by:

\[ \eta_m = \kappa_{m_{(Q)}} - \kappa_{m_{(P)}}. \]  

(12)

The corresponding jump risk premium in \( m \) is the difference between the mean of the jump sizes

\(^2\)Pan (2002) argues that introducing different intensities of jumps under the historical and pricing measure introduces a jump-timing risk premium that is very difficult to disentangle from the mean jump risk premium. The consequence of this assumption is that the jump-timing risk premium is artificially incorporated into the mean jump size risk premium.

\(^3\)In the literature \( \sigma_Y \) has sometimes been constrained to be the same under \( \mathbb{P} \) and \( \mathbb{Q} \) (Bates (1988), Naik and Lee (1990)), but this is not required by absence of arbitrage and we follow Broadie, Chernov, and Johannes (2007) by allowing them to be different. Indeed, they find strong evidence for them to be different and report that this has strong implications for the magnitude of the premium attached to the mean price jump size.
in $m$ under $\mathbb{Q}$ and $\mathbb{P}$. The two latter risk premia should be identified by the cross section and term structure of the VIX implied volatilities as well as the long-term SPX implied volatilities. Therefore, introducing options in our dataset with various moneynesses and maturities is crucial to have meaningful values for these premia.

Finally, no-arbitrage considerations force the volatility of volatilities ($\sigma_v$ and $\sigma_m$) and the correlation between the returns and volatility $\rho_{Y,v}$ to be equal under $\mathbb{P}$ and $\mathbb{Q}$.

### 4.3 Derivatives pricing

Due to the affine property of the VIX$^2$ and given the results of Duffie, Pan, and Singleton (2000), we have the following result:

**Proposition 4.2.** *The Laplace transforms of the returns and VIX$^2$ defined by the model (1) - (3) are given by*

\[
\Psi_{VIX^2}(t, v_t, m_t; \omega) := E^Q_t \left[ e^{\omega VIX^2_T} \right],
\]

\[
\Psi_Y(t, y_t, v_t, m_t; \omega) := E^Q_t \left[ e^{\omega Y_T} \right],
\]

*are exponential affine in the factor processes:*

\[
\Psi_{VIX^2}(t, v, m; \omega) = e^{\alpha(T-t) + \beta(T-t)v + \gamma(T-t)m},
\]

\[
\Psi_Y(t, y, v, m; \omega) = e^{\alpha_Y(T-t) + \beta_Y(T-t)y + \gamma_Y(T-t)v + \delta_Y(T-t)m},
\]

*where $\alpha$, $\beta$, $\gamma$, $\alpha_Y$, $\beta_Y$, $\gamma_Y$ and $\delta_Y$ are functions defined on $[0, T]$ by ODEs presented in the online Appendix B. $\omega \in \mathbb{C}$ is chosen so that the expectations above are well defined.*

In affine models, option pricing is most efficiently performed using Fourier inversion techniques since we know the Fourier transform of the stochastic processes of interest. To price options on the S&P 500, Fang and Oosterlee (2008) report that the Fourier Cosine Expansion is very efficient and fast compared to other Fourier inversion techniques. We use this method to price S&P 500 options and extend it to incorporate also the pricing of VIX options. This technique is comparable to the inversion performed by Sepp (2008a) but more parsimonious in the number of computational parameters.
Pricing options on the VIX poses technical difficulties that are not encountered when pricing equity options. To understand why it is different, let us write the price of a call option with strike $K$ and maturity $T$ on the VIX at time $t = 0$ in the following form:

$$C(VIX_0, K, T) = e^{-rT} \int_0^\infty (\sqrt{v} - K)^+ f_{VIX_T^2}(v) dv,$$

(13)

where $f_{VIX_T^2}$ is the density of the VIX$^2$ at time $t = T$. We introduce the density of VIX$^2$ because this is the variable which is affine in our framework (as opposed to working with the VIX).

The square root appearing in the integral as part of the payoff prevents us from using the Fast Fourier Transform of Carr and Madan (1999). For S&P 500 call options the payoff can be written as $(e^y - K)^+$ where $y$ is the log of the stock price. The fact that we have the exponential $e^y$ allows to interpret this integral as a Fourier transform. To apply the same methodology in the case of VIX derivatives, we would need the log of the VIX to be affine which is incompatible with affine models. This justifies our choice to depart from the standard Fourier pricing techniques.

The basic idea of the method developed by Fang and Oosterlee (2008) is to write the density of the S&P 500 log-returns as a Fourier cosine expansion on a well chosen truncated interval $[a, b]$. This allows them to derive the price of S&P 500 options; we use the same methodology to calculate the price of VIX options.

**Theorem 4.1.** Let us consider a European style contingent claim on the VIX index with maturity $T$ and payoff $u_{VIX}(VIX^2) = (\sqrt{VIX^2} - K)^+$ (respectively on the normalized S&P 500 forward $\tilde{Y} := \log (F/K)$ with payoff $u_{SPX}(e^\tilde{y}) = K(e^\tilde{y} - 1)^+$ at $T$). Given a chosen interval $[a_{VIX}, b_{VIX}]$ for the support of the $VIX^2\big|_{VIX_0, m_0}$ density (respectively $[a_Y, b_Y]$ for the support of the density of $\tilde{Y}_T\big|_{Y_0} = \log (F_{t_0}(T)/K)|_{Y_0}$, the price $P_{VIX}(t_0, VIX_0)$ at time $t = t_0 \geq 0$ (respectively $P_{SPX}(t_0, Y_0)$) of the contingent claim is

$$P_{VIX}(t_0, VIX_0) = e^{-r(T-t_0)} \sum_{n=0}^{N-1} A_n^{VIX^2} U_n^{VIX^2},$$

(14)

respectively:

$$P_{SPX}(t_0, Y_0) = e^{-r(T-t_0)} \sum_{n=0}^{N-1} A_n^Y U_n^Y,$$

(15)

where the prime superscript in the sum $\sum'$ means that the first term $A_0U_0$ is divided by 2. The
terms in the sum are defined by:

\[ A_n^{\text{VIX}^2} = \frac{2}{b_{\text{VIX}} - a_{\text{VIX}}} \text{Re} \left\{ \Psi_{\text{VIX}^2} \left( t_0, v_0, m_0; \frac{n\pi}{b_{\text{VIX}} - a_{\text{VIX}}} \right) \exp \left( -ia_{\text{VIX}} \frac{n\pi}{b_{\text{VIX}} - a_{\text{VIX}}} \right) \right\}, \tag{16} \]

\[ U_n^{\text{VIX}^2} = \int_{a_{\text{VIX}}}^{b_{\text{VIX}}} u_{\text{VIX}}(v) \cos \left( n\pi \frac{v - a_{\text{VIX}}}{b_{\text{VIX}} - a_{\text{VIX}}} \right) dv, \tag{17} \]

respectively

\[ A_n^{\tilde{Y}^2} = \frac{2}{\tilde{b}_{\tilde{Y}} - a_{\tilde{Y}}} \text{Re} \left\{ \Psi_{\tilde{Y}^2} \left( t_0, \tilde{Y}_0, v_0, m_0; \frac{n\pi}{\tilde{b}_{\tilde{Y}} - a_{\tilde{Y}}} \right) \exp \left( -ia_{\tilde{Y}} \frac{n\pi}{\tilde{b}_{\tilde{Y}} - a_{\tilde{Y}}} \right) \right\}, \tag{18} \]

\[ U_n^{\tilde{Y}^2} = \int_{a_{\tilde{Y}}}^{\tilde{b}_{\tilde{Y}}} u_{\text{SPX}}(e^{\tilde{y}}) \cos \left( n\pi \frac{\tilde{y} - a_{\tilde{Y}}}{\tilde{b}_{\tilde{Y}} - a_{\tilde{Y}}} \right) d\tilde{y}, \tag{19} \]

where \( \Psi_{\tilde{Y}^2} \) is the Laplace transform of \( \tilde{Y}_T = Y_T - \log K \), i.e.,

\[ \Psi_{\tilde{Y}^2}(t, \tilde{y}, v, m; \omega) = \Psi_{Y_T}(t, y, v, m; \omega)e^{-\omega \log K}. \]

We note that the coefficients \( A_n^{\text{VIX}^2} \) and \( A_n^{\tilde{Y}^2} \) are computed using Proposition 4.2 and the coefficient \( U_n^{\text{VIX}^2} \) is known in closed form and given in Appendix B. Finally, a closed form for \( U_n^{\tilde{Y}^2} \) can be found in Fang and Oosterlee (2008).

5 Joint estimation and particle filter

The goal of this section is twofold. First, we explain how we calibrate the nested models (1) - (3) to S&P 500 and VIX options, i.e., we estimate the model under the pricing measure using the VIX and S&P 500 option price surfaces on a given date. This exercise allows us to show that the \( Q \) dynamics of the model is sufficiently rich to accurately price both S&P 500 and VIX derivatives together, i.e., at any date \( t \) we can find a fixed set of parameters which allows the model to price both VIX options and S&P 500 options accurately. Second and most importantly, we detail how we have built a time consistent estimation of the models using a time series of S&P 500 and VIX indices together with a time series of S&P 500 and VIX option prices. This means that we estimate both the \( P \) and \( Q \) dynamics of the model using the time series of indices and options (i.e., we find one vector of parameters for the whole time series of SPX/VIX spots and SPX/VIX options). The algorithm we use is the Auxiliary Particle Filter, introduced by Pitt and Shephard (1999). It allows to filter out
unobserved latent variables, such as the volatility process or jumps.

From Section 4.1, we recall the $\mathbb{P}$- and $\mathbb{Q}$- parameter vectors:

\[
\Theta^\mathbb{P} = \{\kappa^{\mathbb{P}}_v, \kappa^{\mathbb{P}}_m, \theta^{\mathbb{P}}_m, \nu^{\mathbb{P}}_v, \nu^{\mathbb{P}}_m, \mu^{\mathbb{P}}_Y, \sigma^{\mathbb{P}}_Y, \eta_Y\} \\
\Theta^\mathbb{Q} = \{\kappa^{\mathbb{Q}}_v, \kappa^{\mathbb{Q}}_m, \theta^{\mathbb{Q}}_m, \nu^{\mathbb{Q}}_v, \nu^{\mathbb{Q}}_m, \mu^{\mathbb{Q}}_Y, \sigma^{\mathbb{Q}}_Y\}.
\]

The remaining parameters are equal under both measures:

\[
\Theta^{\mathbb{P},\mathbb{Q}} = \{\lambda^Y_v, \lambda^Y_0, \lambda^Y_1, \lambda^m_m, \lambda^m_{12}, \lambda^m_0, \lambda^m_{10}, \sigma_m, \sigma_v, \rho_{Yv}\}.
\]

The vector of all parameters is then $\Theta = \{\Theta^\mathbb{P}, \Theta^\mathbb{Q}, \Theta^{\mathbb{P},\mathbb{Q}}\}$.

5.1 Least-squares calibration - Methodology

In this approach, we calibrate our model to the cross section of S&P 500 and VIX options on some chosen dates. On each date, the output will be a set of values for the risk-neutral parameters $\Theta^\mathbb{Q}$ and $\Theta^{\mathbb{P},\mathbb{Q}}$. Calibration to one single day of options data does not allow us to estimate the parameters $\Theta^\mathbb{P}$ since options are priced under the pricing measure $\mathbb{Q}$ and no time series is used. This exercise is important because if the model is not able to reproduce well the implied volatility patterns of both markets together on a single date, then there is no point in estimating the model using a filter on a time series of options and indices.

We fix a date $t$. Let us consider $\{IV_{SPX_i}^{Mkt}\}_{i \in I}$ the set of implied volatilities of options on the S&P 500 for the strikes $\{K_i\}$ and maturities $\{T_i\}$ available in our dataset\footnote{The dataset is described in the empirical analysis section 3 where we explain how implied volatilities have been calculated from S&P 500 and VIX options.} for this date. We use the superscript $Mkt$ for 'Market' implied values. We denote by $\{IV_{VIX_j}^{Mkt}\}_{j \in J}$ the set of VIX option implied volatilities on the same date $t$. $I = \{1, ..., \#I\}$ is the set of integers indexing S&P 500 options available for this date and $J$ the set indexing VIX options. To estimate parameter values, we minimize the distance between market and model implied volatilities (or option prices equivalently).

We have chosen two distance criteria\footnote{Since we analyze the fits in section 6 in terms of implied volatilities and not option prices, we do not consider other popular choices of distances including absolute error of the logarithm of option prices, relative error of option prices (see Christoffersen and Jacobs (2004)) . Alternatively, we checked that using distances taking into account the bid-ask} that put different emphasis on S&P 500 and VIX options as...
well as on at-the-money (ATM) and out-of-the-money (OTM) options. We denote by $IV_{SPX_i}^{Mod}$ the model implied volatility of option with strike $K_i$ and maturity $T_i$ (respectively $IV_{VIX_j}^{Mod}$ corresponding to the notations above). The root mean squared error (RMSE) in implied volatilities on date $t$ is defined as:

$$RMSE_{SPX}(t) := \sqrt{\frac{1}{\#I} \sum_{i \in I} (IV_{SPX_i}^{Mkt} - IV_{SPX_i}^{Mod})^2}$$

$$RMSE_{VIX}(t) := \sqrt{\frac{1}{\#J} \sum_{j \in J} (IV_{VIX_j}^{Mkt} - IV_{VIX_j}^{Mod})^2}$$

$$RMSE(t) := \frac{1}{2} (RMSE_{SPX}(t) + RMSE_{VIX}(t))$$ (23)

We furthermore consider the average relative error (ARE) in implied volatilities on date $t$:

$$ARE_{SPX}(t) := \frac{1}{\#I} \sum_{i \in I} \frac{|IV_{SPX_i}^{Mkt} - IV_{SPX_i}^{Mod}|}{IV_{SPX_i}^{Mkt}}$$

$$ARE_{VIX}(t) := \frac{1}{\#J} \sum_{j \in J} \frac{|IV_{VIX_j}^{Mkt} - IV_{VIX_j}^{Mod}|}{IV_{VIX_j}^{Mkt}}$$

$$ARE(t) := \frac{1}{2} (ARE_{SPX}(t) + ARE_{VIX}(t)) .$$ (24)

Since the IVs are the highest for OTM SPX puts and OTM VIX calls, the RMSE puts more emphasis on fitting these options (which are the most liquid together with ATM options). On the other hand, it is also arguable that a 1% absolute error on an IV does not have the same importance if the market IV is 10% or 80%. The average relative error distance ARE takes this consideration into account by computing relative errors.

To cope with the ill-posedness of the calibration problem and the potential existence of multiple minima, we use two global optimizers namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), introduced by Hansen and Ostermeier (1996), and the Differential Evolution (DE) algorithm introduced by Storn (1996). They are evolutionary algorithms designed for high-dimensional non-linear non-convex optimization problems in a continuous domain. They are based on the principle of biological evolution, i.e., at every step new vectors of parameters are generated based on spread of IVs as in Cont and Kokholm (2011) does not significantly change the quality of fits.

---

6We are grateful to Jochen Krause for his implementation of various evolution optimizers including the CMA-ES and DE algorithms.
the optimal set of parameters up to that step and random perturbations, the objective function is evaluated for each of these new parameter vectors, and the new optimal parameter set becomes the one which minimizes the objective function.

5.2 Particle filter

While the least-square calibration provides us with a static estimation of parameter values, it is more insightful to use the whole time series of option prices and index levels to learn about the dynamic properties of the unobservable processes (volatility, central tendency, jumps) and the risk premia associated to them. Sequential Monte-Carlo methods are ideal for this purpose as they allow to sequentially filter the trajectories of latent processes based on the information available. As they take as input a time series of observations, they furthermore allow to better identify parameters and therefore deliver more robust estimates. We discretize the continuous-time model under $\mathbb{P}$ on a uniform grid composed of $M + 1$ points $t \in \{ t_0 = 0, t_1 = \Delta t, \ldots, t_k = k\Delta t, \ldots, t_M = M\Delta t \}$ (for some $M \in \mathbb{N}^*$). For any $t = t_k$, (0 ≤ $k$ ≤ $M - 1$) we obtain the corresponding state-space discretization:

$$
\Delta Y_t = Y_{t+\Delta t} - Y_t = \left[ -\lambda^Y(v_t, m_t)(\theta_{Z_t}^{(P)}(1, 0, 0) - 1) - \frac{1}{2}v_t + \gamma_t \right] \Delta t + \sqrt{v_t} \Delta W_t^{Y_t}
$$

$$+
Z_t^{Y_t} \Delta N_t^{Y_t} \tag{25}
$$

$$
\Delta v_t = v_{t+\Delta t} - v_t = \kappa_v^{(P)} \left( \frac{\kappa_v^{(Q)}}{\kappa_v^{(P)}} m_t - v_t \right) \Delta t + \sigma_v \sqrt{v_t} \Delta W_t^{v_t} + Z_t^{v_t} \Delta N_t^{Y_t} \tag{26}
$$

$$
\Delta m_t = m_{t+\Delta t} - m_t = \kappa_m^{(P)} (\theta_m^{(P)} - m_{t-}) \Delta t + \sigma_m \sqrt{m_t-} \Delta W_t^{m_t} + Z_t^{m_t} \Delta N_t^{m_t}. \tag{27}
$$

In practice, $\Delta t$ will correspond to one day, since we use daily data. In particular, we do not augment the time space since Johannes, Polson, and Stroud (2009) have shown that the advantage of introducing additional time steps is very limited when using daily observations.

The latent factors we wish to infer from the observations is: $L_t = \{ v_t, m_t, \Delta N_t^{Y_t}, \Delta N_t^{m_t}, Z_t^{Y_t}, Z_t^{v_t}, Z_t^{m_t} \}$. Note that among these factors, only $v_t$ and $m_t$ depend on their past values, as the jump sizes $(Z_t^{Y_t}, Z_t^{v_t}, Z_t^{m_t})$ are i.i.d. over time and so are the increments of Poisson processes conditionally on $v_t$ and $m_t$. Equation (25) is the first measurement equation. The second one is given by the observation of the VIX index level with error. Indeed, since the VIX index is in practice calculated using a finite number of options, a discretization bias is introduced. Furthermore, Jiang and Tian
(2007) point to systematic biases in the VIX. We write this error as follows:

\[ \text{VIX}_t^2 - (\alpha_{\text{VIX}^2} v_t + \beta_{\text{VIX}^2} m_t + \gamma_{\text{VIX}^2}) = \epsilon_t^{\text{VIX}}. \] (28)

The error terms \( \epsilon_t^{\text{VIX}} \) are assumed to follow a normal distribution with mean zero and variance \( s > 0 \). Other observable quantities are the prices of S&P 500 and VIX options. We assume that option prices are observed with an error, which is due to different sources such as the bid-ask spread, the processing and timing errors (all options considered on a given day are not traded at the same time) and misspecification error. The observation equations (29) - (30) for options are:

\[ \frac{O_{t,i}^{\text{SPX, Mod}}(Y_t, v_t, m_t, \Theta^Q, \Theta^P, Q) - O_{t,i}^{\text{SPX, Mkt}}}{O_{t,i}^{\text{SPX, Mkt}}} = \epsilon_{t,i}^{\text{SPX, options}} \] (29)

\[ \frac{C_{t,j}^{\text{VIX, Mod}}(v_t, m_t, \Theta^Q, \Theta^P, Q) - C_{t,j}^{\text{VIX, Mkt}}}{C_{t,j}^{\text{VIX, Mkt}}} = \epsilon_{t,j}^{\text{VIX, options}} \] (30)

where \( O_{t,i}^{\text{SPX, Mkt}} \) corresponds to the market price at time \( t \) of the S&P 500 option indexed by \( i \in I \), \( O_{t,i}^{\text{SPX, Mod}}(Y_t, v_t, m_t, \Theta^Q, \Theta^P, Q) \) to the model price of the same option assuming the \( Q \) parameters are \{\( \Theta^Q, \Theta^P, Q \)\}. Similarly, \( C_{t,j}^{\text{VIX, Mkt}} \) denotes the market price at time \( t \) of the call option on the VIX indexed by \( j \in J \), and \( C_{t,j}^{\text{VIX, Mod}}(v_t, m_t, \Theta^Q, \Theta^P, Q) \) the model price with \( Q \) parameters \{\( \Theta^Q, \Theta^P, Q \)\}.

We assume the error terms to be normally distributed and heteroscedastic:

\[ \epsilon_{t,i}^{\text{SPX, options}} \sim N(0, \sigma_{\epsilon_{t,i}^{\text{SPX}}}^2), \] (31)

and

\[ \epsilon_{t,j}^{\text{VIX, options}} \sim N(0, \mu_{t}^{\text{VIX}}, \sigma_{\epsilon_{t,j}^{\text{VIX}}}^2), \] (32)

where \( \mu_{t}^{\text{VIX}} \) is proportional to the error \( \epsilon_t^{\text{VIX}} \) which has been made on the estimation of the VIX level. Indeed, if the underlying’s value is not accurately estimated, it introduces a systematic bias on the calculation of VIX option prices. We specify the variance of errors as follows:

\[ \sigma_{\epsilon_{t,i}^{\text{SPX}}}^2 = \exp \left( \phi_0 \text{bid-ask spread}_i + \phi_1 \log \left( \frac{K_i}{F_{t}^{\text{SPX}}(T_i)} \right) \right) + \phi_2(T_i - t) + \phi_3 \] (33)
\[ \sigma_{V_t}^2 = \exp \left( \psi_0 \text{bid-ask spread}_j + \psi_1 \log \left( \frac{K_j}{F_{t,VIX}(T_j)} \right) + \psi_2(T_j - t) + \psi_3 \right). \] (34)

with \( \phi_i \) and \( \psi_i \) are in \( \mathbb{R} \), \( i \in \{0, \ldots, 3\} \).

At time \( t = t_k \) (for \( 0 \leq k \leq M \)), we denote by \( y_t \), the set of observable prices. The log-likelihood of a time-series of \( n + 1 \) observations \( y^{tn} = (y_{t_0}, \ldots, y_{t_n}) \) (\( n \leq M \)) with joint density \( p \) conditionally on a set of parameters \( \Theta \) and a model specification \( M \) is equal to:

\[ \log L(y^{tn}|\Theta, M) = \sum_{k=1}^{n} \log p(y_{t_k}|y^{tk-1}, \Theta, M) + \log p(y_{t_0}|\Theta, M) \] (35)

where

\[ p(y_{t_k}|y^{tk-1}, \Theta, M) = \int p(y_{t_k}|L_{t_k}, \Theta, M)p(L_{t_k}|y^{tk-1}, \Theta, M)dL_{t_k}. \] (36)

Given an initial density \( p(L_{t_0}|\Theta, M) \), the transition density of state variables \( p(L_{t_k}|L_{t_{k-1}}, \Theta, M) \) and the likelihood function \( p(y_{t_k}|L_{t_k}, \Theta, M) \), filtering methods allow to estimate the distribution \( p(L_{t_k}|y^{tk}, \Theta, M) \) of the current state at time \( t_k \) given all observations up to that time. In particular, particle filters are perfectly adapted to our problem since they can handle nonlinear systems with non-Gaussian innovations. The key idea is to approximate the posterior density function of the latent variables \( p(L_{t_k}|y^{tk}, \Theta) \) by a sum of point masses positioned at strategic points called particles \( \{L_{t_k}^{(i)}\}_{1 \leq i \leq n_p} \):

\[ \hat{p}(L_{t_k}|y^{tk}, \Theta, M) = \sum_{i=1}^{n_p} \pi_{t_k}^{(i)} \delta(L_{t_k} - L_{t_k}^{(i)}) \] (37)

where \( \pi_{t_k}^{(i)} \) denotes the normalized importance weight for particle \( i \) and \( \delta(.) \) is the Dirac function. \( n_p \) is the number of support points (particles) for \( \hat{p}(L_{t_k}|y^{tk}, \Theta, M) \).

To apply the particle filter, one needs to be able to simulate at every time \( t_k \) a number \( n_p \) of particles \( L_{t_k}^{(i)}, i = 1, \ldots, n_p \) from \( p(L_{t_k}|y^{tk-1}, \Theta, M) \) and to be able to evaluate \( p(y_{t_k}|L_{t_k}^{(i)}, \Theta, M) \). Based on these simulated particles, \( p(y_{t_k}|y^{tk-1}, \Theta, M) \) is approximated by:
Multiple versions of the particle filter exist. We use the Auxiliary Particle Filter (APF) proposed by Pitt and Shephard (1999) and extend the approach described in Johannes, Polson, and Stroud (2009). The main advantage of the APF compared to more basic particle filters such as the Sampling Importance Resampling (SIR) is that it is more capable of detecting jumps compared to the SIR filter which faces sample impoverishment leading to particle degeneracy. Both filters are described in Johannes, Polson, and Stroud (2009) for filtering latent factors from returns in a Heston model with jumps. The authors also use it with option prices but do not provide details on the adjustments made. Our extension of the filter makes it possible to extract the most probable paths of both factors \( v \) and \( m \) as well as the jump components from the set of observable variables \( y^n \). To incorporate the information contained in the S&P 500 and VIX levels, we calculate at every time \( t \) the joint probability of having 0 or 1 jump in every process given the new observations. Since the jump size of returns is normally distributed but jumps in the variance processes are exponentially distributed, the conditional likelihood of the new observations given a combination of jumps involves the sum of a normal and (up to two) exponential random jumps. To compute the joint probability of jumps and preserve tractability, we approximate the exponentially distributed jump sizes by a categorical distribution (generalization of a Bernoulli distribution) which has support a certain number of chosen quantiles. Finally, to overcome the usual difficulty of particle filters to recognize jumps, we systematically simulate jumps in one tenth of the particles.8

The detailed filtering procedure is described in Appendix D.

6 Daily calibration results

The first step in evaluating the performance of the model (1)-(3) is to calibrate it to one day of options data to make sure that the model is flexible enough to simultaneously price options on both markets.

---

7 Robustness tests were performed on simulated data to check that the choice of quantiles was appropriate.

8 We tested the accuracy of the filter using different proportions of particles with systematic jumps and found that 1/10 was a reasonable choice.
We follow the method outlined in section 5.1. We have chosen some dates on which we calibrate the model (1)-(3) to the cross section of S&P 500 and VIX implied volatilities. We only report the results on four dates as they are representative of the whole sample. We consider two days where the market was facing great uncertainty about the future, October 22 2008 (one month after the bankruptcy of Lehman Brothers) and May 05 2010, at the beginning of the European sovereign debt crisis. On these dates, the markets were under stress and S&P 500 implied volatilities had a very strong negative skew and levels above 100% for short term options. The other two days we report are rather calm compared to those: July 11 2007 and June 10 2009.

Gatheral (2008) has shown that the Heston model is incapable of reproducing the positive skew in VIX implied volatilities (IVs) as displayed in Figure 4 and Sepp (2008a,b) added that introducing positive jumps in the volatility dynamics of the Heston model allows the model to have a positive skew in VIX IVs. As a consequence, departing from the usual literature on S&P 500 option pricing, the simplest model we consider is the Heston model with jumps in returns and volatility. It corresponds to the model (1)-(2) where the central tendency \( m \) is constant. We will denote this model by SVJ (Stochastic Volatility with Jumps). The most flexible model we consider, with 2 factors to represent the volatility component namely the variance and a stochastic central tendency, is referred to as SVJ2.

We report in Table 3 the results for the RMSE calibration with respect to implied volatilities (23).\(^9\) We emphasize that for each model and on each day, we have minimized the total RMSE from the VIX and the S&P 500 market together. We report the resulting \( \text{RMSE}_{SPX} \) and \( \text{RMSE}_{VIX} \) since these are indicative of the quality of the fit on each market. Irrespective of the day, we observe that the SVJ and SVJ2 models perform comparably on the S&P 500 options market, both fitting very well with an average RMSE of around 1.5% across the dates we have calibrated to. In contrast, we see that there are dates when the SVJ model struggles to fit the VIX IVs in addition to the S&P 500 IVs whereas the SVJ2 model satisfactorily fits both. This is observed for instance on July 11 2007 and on May 05 2010. We can see the comparative fits for the VIX options market in Figure 5.\(^10\) Panels A to D correspond to the SVJ2 model’s fit and panels E to H to the SVJ model’s fit. On

---

\(^9\)The results are qualitatively similar when minimizing over the distances (23) and (24), we therefore only report one result.

\(^10\)The SVJ and SVJ2 both match the S&P 500 options market prices almost perfectly so we do not show them.
this date, it seems that for short maturities, the SVJ model has difficulties to reproduce the strong positive skewness of VIX IVs (which was already the case for the Heston model). This shortcoming of the SVJ is not often noticeable and we therefore do not make it a general statement. Indeed, on the other two dates we report, the fits of the SVJ and SVJ2 are comparable on the VIX market.

[Insert Table 3 here]

[Insert Figure 5 here]

Daily calibration is essentially a multiple curve fitting exercise, where we check whether the models can fit the risk-neutral distributions inferred by option prices at different maturity. A more thorough analysis is needed to conclude at this point that the SVJ2 is much better than the SVJ model to price S&P 500 and VIX options together. Indeed the SVJ2 has more parameters and is therefore bound to fit better. Furthermore, some of the parameters we get from daily calibrations can vary a lot from one day to the next.\textsuperscript{11} At this point, it is therefore not possible to know whether the dynamics of the model (1)-(3) can reproduce the time evolution of these smiles. This is what we will focus in the next section.

7 Time-series estimation results using a Particle Filter

The second step in evaluating the performance of the model is to estimate it using the time series of S&P 500 and VIX indices together with S&P 500 and VIX options. This is achieved using the particle filter described in Section 5.2. We report the results for different sub-models of (1)-(3) and analyze the gain of information and robustness we have from adding the data from the VIX market to the dataset. The sub-models considered of the SVJ2 model are the SVJ model and the 2-factor continuous model, which has a stochastic central tendency but no jumps. We refer to the latter model as SV2.

\textsuperscript{11}As explained in Broadie, Chernov, and Johannes (2007) and Lindström, Ströjby, Brodén, Wiktorsson, and Holst (2008), the parameters obtained when calibrating to daily options prices are not stable over time. To better understand how the model performs over time, it is important to estimate the model on a time series of options data.
7.1 Specific data treatment for the Particle filter

The dataset contains a large amount of ATM options compared to OTM and deep OTM options. This implies that if we use the filter (inside the maximum likelihood procedure) on this dataset and the model is not able to fit all options, the fitting of ATM options will be its priority rather than deep OTM options. Given the formula in Breeden and Litzenberger (1978), this results in fitting the body of the S&P 500 returns distribution rather than the tails which is not what we want: We need information about extreme events contained in the data to be incorporated into the models. For this reason, we have decided to interpolate the S&P 500 slices and to re-sample option prices from our parametric fit uniformly across moneynesses.\(^{12}\) Other advantages of our interpolating method is that the resulting data is arbitrage free,\(^{13}\) we have fewer points for each slice (but still representing accurately the information of each slice), thus reducing computational complexity.

We explain in detail in Appendix C how we have used the efficient mixture of log-normals approach of Rebonato and Cardoso (2004) to have a parametric fit for each S&P 500 implied volatility slice. The root mean squared error of the S&P 500 implied volatilities parametric fits are on average around 0.25\% and we therefore do not loose information especially given the market bid-ask spread. Finally, given the parametric fit for a given slice, we sample a fix number (we have chosen 15) of option prices uniformly distributed on the moneyness axis.

We do not perform any interpolation for the VIX options dataset as most VIX options are OTM and therefore contain information about the tails of the VIX distribution (i.e., variance and central tendency processes).

We divide the data into four different datasets:

*Dataset 1:* S&P 500 returns and VIX index levels,

*Dataset 2:* S&P 500 returns, VIX index levels and S&P 500 options,

*Dataset 3:* S&P 500 returns, VIX index levels and VIX options,

*Dataset 4:* S&P 500 returns, VIX index levels, S&P 500 options and VIX options.

\(^{12}\)It is common to interpolate data, see, e.g., Broadie, Chernov, and Johannes (2007). This eliminates arbitrage opportunities in the data and removes the accumulation of options around the ATM region.

\(^{13}\)Since we have considered mid-prices and because of synchronization issues between the underlying and the options, implied volatility slices are a priori not guaranteed to be arbitrage free.
Using these different datasets allows us to make inference on what information they contain, and whether they are consistent with one another. Our filter uses daily time steps and incorporates information on the underlying indices on a daily basis. As the database comprises a large amount of options, it is unfeasible to calculate option prices every day for every particle, we follow Pan (2002) and Johannes, Polson, and Stroud (2009) among others and use weekly (Wednesday) option data. Furthermore, this eliminates beginning-of-week and end-of-week effects.

The time-series of observations is decomposed into two periods, the first one from March 1st, 2006 to October 10, 2008 (before the peak of crisis and the VIX index increased to its highest point). This is a rather calm period,\textsuperscript{14} that we will use as in-sample estimation period. Our out-of-sample period starts on November 25, 2008 and ends on October 29, 2010. This period includes very high levels of volatility (i.e., implied volatilities from S&P 500 and VIX options as well as VIX index values). The last column of Table 4 reports the amount of options within each moneyness and maturity range in both periods.

\[\text{Insert Table 4 here}\]

In particular, our dataset contains 608 close-to-maturity OTM call options on the S&P 500 with moneyness larger than 1.05 and 2243 OTM put options with moneyness smaller than 0.95 in the in-sample period. These options have maturity shorter than two months. Analogously in the out-of-sample period the dataset comprises 737 close-to-maturity OTM call and 2032 OTM put options. As highlighted in Bollerslev and Todorov (2011), these options provide valuable information on jumps as they have small value unless a large movement in the S&P 500 occurs. Similarly, OTM calls on the VIX with short maturity contain information on the extreme upwards moves in the VIX index, and help identify the heavy-tailedness of the right tail of the VIX distribution. Our dataset comprises 1006 such options on the VIX with moneyness larger than 1.1 in the in-sample period and 1269 in the out-of-sample period.

\textsuperscript{14}We have decided to include the beginning of October 2008 so that the in-sample period actually includes several dates with extreme events.
7.2 Filtered trajectories

Figure 7 displays the filtered trajectory of the volatility process when estimating all three models SVJ, SV2 and SVJ2 over Dataset 4. Although the volatility trajectories are consistent across models, we notice that the volatility is slightly more variable when the calibration is done with models with a stochastic central tendency. Indeed, the Feller condition is imposed on the SVJ model, which restricts the amplitude of volatility movements. This condition is relaxed for 2-factor models as the long-term mean of the variance is varying, which allows the volatility process to have a larger amplitude. In particular, for all datasets the estimated parameter $\sigma_v$, which controls the volatility of volatility, is considerably smaller for the SVJ model than for the SV2 and SVJ2 models, see Table 5. Figure 8 represents the difference between the filtered volatility processes using Dataset 4 and the the other datasets (Datasets 1 to 3). Until the peak in the VIX toward the end of the in-sample period, this difference is very small (lower than 2%). In this period, the filtered volatility using Dataset 1 remains slightly lower than the volatility filtered using the other datasets so options seem to contribute to increase the filtered volatility only moderately. When the VIX index reaches its highest peak at the end of 2008, the volatility filtered using Dataset 4 is on average close to the one filtered using Dataset 3 but up to 23% lower than the volatility filtered using Datasets 1 and 2. It appears that adding S&P 500 options to Dataset 1 does not provide much new information on the behavior of the variance. Conversely, adding VIX options brings information on the volatility in times of market turmoil which is not spanned by the underlying levels and the S&P 500 options. In the out-of-sample period, the difference between the trajectories remains within ±3% except during the second smaller peak of variance in May 2010. Figure 9 shows the jump sizes when then the filtered jump probability is larger than 50%. Small jumps (around 2%) are filtered in the variance process at the beginning of 2007, and larger jumps (above 10%) are filtered towards the end of 2008 as the VIX peaks, when VIX options are not part of the dataset. Using Dataset 3, the algorithm only filters a small jump in $v$ at the beginning of 2007, but finds a large jump likely to occur in the trajectory of $m$ in November 2008, see Figure 10.

[Insert Figures 7, 8, 9 and 10 here]

[Insert Table 5 here]

The process $m$ exhibits the same general shape across estimation datasets and 2-factor models, see
Figure 11. However, we notice that the central tendency process is much more reactive to information in options’ markets than to indices and that its magnitude increases when options are part of the estimation dataset, especially during crises. Using *Dataset 1*, $m$ reaches a maximum lower than 15% in 2009 (see Figure 15), but it can go almost as high as 35% when S&P 500 or VIX options are added to the dataset. The trajectories filtered by the SV2 and SVJ2 models almost overlap before the crisis, however the filtered $\{m_t\}_{t\geq 0}$ increases earlier for the SVJ2 model than with the SV2 model, suggesting that including jumps in the model allows to be more reactive and adjust the central tendency faster. The process $m$ is overall more stable and less erratic than the variance process, giving evidence that it captures long-term trends. In fact, its volatility parameter $\sigma_m$ is in the range $[10\%, 30\%]$ for both the SV2 and SVJ2 models regardless of the dataset chosen for the estimation. Its speed of mean-reversion is more than three times smaller than that of $v$ under $P$ and more than six times smaller under $Q$. The process is therefore more persistent. While the variance process increases dramatically during the crisis but then returns to a level which is comparable to the one before the crisis, the central tendency also increases during the crisis but then goes down to a level which is higher than before the crisis (between 5% and 10% after the crisis, compared to 3% before). Therefore $v$ captures the punctual movements of the variance while $m$, being the stochastic long-term reversion level of $v$, embeds longer-term expectations of investors regarding the variance and can therefore be seen as an indicator of market turmoil. Even though the process $m$ is more stable than $v$, the approximation made by the SVJ model of the process $m$ being a constant is too rough. The level which the central tendency reaches during and after the crisis (especially in *Datasets 2,3,4* that include options) is clearly underestimated by the SVJ model. In fact, the constant central tendency estimated in the SVJ model seems to be close to the average filtered central tendency of the SV2 and SVJ2 models over the in-sample period. This makes the SVJ model insensitive and non-adaptable to different regimes in long-term volatility especially when the in-sample estimation period exhibits more instabilities than the in-sample period.

[Insert Figure 11 here]

### 7.3 Are jumps and/or a stochastic central tendency needed?

The filtered trajectories of the central tendency $m$ for models SV2 and SVJ2 show that this process is clearly not constant. However, it could be the case that having $m$ stochastic has a limited impact
in terms of pricing and forecasting performances. In this section, we precisely analyze whether jumps and a stochastic central tendency are needed to reproduce the features of VIX levels, S&P 500 option prices and finally VIX option prices.

Let us first investigate whether these features are needed to provide an accurate fit of the VIX index. Given the filtered trajectories for the processes \( v \) and \( m \) inferred by Dataset 1 we calculate the corresponding model-implied values of the VIX index using the optimal parameters. As illustrated by Figure 12, the model accurately reproduces the time-series of VIX index values. Table 6 reports the corresponding Mean Errors (ME) and Root Mean Square Errors (RMSE) and shows that they are comparable across models. This observation is consistent across datasets. At first sight, jumps and a stochastic central tendency therefore seem superfluous to reproduce the trajectory of the VIX level.

To statistically challenge this claim, we run likelihood tests. Table 5 reports the log-likelihood values as well as the values of the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) for Datasets 1 to 4. When focusing on Datasets 1, both criteria are slightly in favor of the SVJ2 model and therefore support the use of jumps and of a stochastic central tendency.

We further challenge the in- and out-of-sample performance of the SVJ2 model by running various Diebold-Mariano (DM) tests. For the three models considered, we consider two loss functions namely the absolute and quadratic pricing errors, respectively defined as \( L(e_t) = |e_t| \) and \( L(e_t) = e_t^2 \), where \( e_t \) refers to the difference at time \( t \) between the model-forecast of the VIX and the true value of the index. We denote the loss differential at time \( t \) between the SVJ or the SV2 model and the SVJ2 model by \( d^{SVJ/SV2,SVJ2}_t = L(e^{SVJ/SV2}_t) - L(e^{SVJ2}_t) \). If the two models considered have comparable pricing errors, then \( \mathbb{E}^P[d^{SVJ/SV2,SVJ2}_t] = 0 \). A positive value of the expectation means that the SVJ2 model outperforms the sub-model considered. Table 7 reports the results when the calibration is done using Dataset 1 and shows that the test values are very close to zero with the quadratic loss function and negative with the absolute loss function, suggesting that the SVJ and SV2 models should be preferred to the SVJ2 model when calibrating them to underlying levels only. When including options
in the estimation dataset, we obtain test values which are very close to zero (within ±0.1 bounds). Therefore we conclude that the three models considered perform comparably at reproducing the trajectories of VIX levels, in- and out-of-sample.

[Insert Table 7 here]

The calibration to S&P 500 options (Dataset 2) highlights the superiority of the SVJ2 and SV2 models over the SVJ model. Computationally, the SV2 model is faster to calibrate as the detection and estimation of jumps involves rare events and is challenging for the particle filter. But more importantly, the SVJ model exhibits Root Mean Square Errors (RMSEs) and Root Mean Square Relative Errors (RMSREs) when pricing S&P 500 options which are for most option categories higher than those of the SV2 model. In particular, the SVJ model does not represent well the deep OTM and long-maturity options, see Tables 4 and 8. The corresponding RMSREs are almost twice those of the 2-factor volatility models for in-sample deep OTM calls and about twice for in-sample long-maturity options. In the out-of-sample period, the poorer performance of the SVJ model extends to medium-maturity options and OTM calls. Therefore a stochastic central tendency allows to better price long-term and deep OTM S&P 500 options. This supports the hypothesis that the process captures the long-term trends of volatility, and therefore enables to better reproduce the term structure of S&P 500 option prices.

[Insert Table 8 here]

The calibration to VIX options (Dataset 3) also favors the SVJ2 model, which yields better RMSEs and RMSREs of VIX option prices than the SVJ and SV2 models except for deep OTM options. The comparison of the SV2 and SVJ models shows that the SVJ model better prices these deep OTM calls while the SV2 model is more appropriate for other moneyness levels. The SVJ model can therefore better represent the tail of the volatility distribution. Consistently with the results we obtained when estimating models to Dataset 2, the SV2 model outperforms the SVJ model in pricing the medium-maturity VIX options (which are the longest maturity VIX options).

When including options on both markets in the estimation dataset (Dataset 4), the SVJ2 model yields RMSEs and RMSREs which are smaller than the SVJ and the SV2 models in-sample in pricing most S&P 500 and VIX option categories, see Tables 4 and 8. We notice that the SVJ model performs
particularly worse at pricing the deep OTM options on the S&P 500, consistently with the results obtained when calibrating the models to *Dataset 2*. Using the Diebold-Mariano test, we investigate whether the SVJ2 model’s pricing performance of SPX and VIX options is significantly better than that of its sub-models on average. For this purpose, we consider two loss functions, the Mean Square Error (MSE) of option prices and the Mean Square Relative Error (MSRE). The resulting test values are presented in Tables 10 and 9. They confirm that the SVJ2 model provides significantly better in-sample MSREs for S&P 500 options than the two other models. As this is not the case when calibrating the model to *Dataset 2*, these tests suggest that the flexibility of the SVJ2 model allows reaching smaller average pricing errors on S&P 500 options only when options on both markets are included in the estimation. The test values associated to VIX option pricing are essentially positive. We however note that in the out-of-sample period, the SV2 model yields smaller relative errors than the SVJ2 model, which might be due to a problem of identification of the jump terms.

Therefore, we conclude that a stochastic central tendency adds significant value in the pricing long-term options and the representation the tails of the returns’ distribution (OTM calls on the S&P 500). On the other side, jumps add value to represent the right tail of the variance distribution (OTM calls on the VIX). Therefore the jumps and the stochastic central tendency of the SVJ2 model provide important improvements over the SVJ and SV2 to represent the underlying returns in a way that is consistent with S&P 500, VIX levels and their derivatives prices. This conclusion is however mitigated by the difficulty to identify jump terms.

Furthermore we observe that the SVJ2 model encounters problems during the crisis and does not well represent volatility smiles. In particular, OTM puts on the S&P 500 tend to be underpriced and OTM calls are generally overpriced, i.e., the smile of volatility does not exhibit enough skewness. This phenomenon affects short-maturity options in particular. Figure 13 compares the moments of S&P 500 returns as implied by market and model option prices, when the models are calibrated to *Dataset 4* (all indices and options). While the skewness of the returns is well represented at the beginning of the in-sample period, it is underestimated from late 2007 until the end of our sample. In the out-of-sample period this phenomenon becomes much more apparent, and all three models yield an implied skewness which is about half the one implied by the market. Similarly, the kurtosis is only slightly underestimated at the beginning of the time-series, but in the out-of-sample period
the model kurtosis it is about half the market implied kurtosis. We add that there is no improvement in the representation of SPX implied moments when adding the options on the VIX market to the estimation dataset.

Figure 14 displays the implied moments of VIX returns and compares the market values to the model values of these moments. From the end of 2007 we notice that the SVJ model yields very variable moments, which are very far from those implied by the market. In particular, the skewness generated by the SVJ is very often negative and the kurtosis reaches very large values exceeding 20. The SVJ2 and SV2 models perform significantly better, however their ability to represent the skewness and kurtosis drastically deteriorates at the very end of the in-sample period. In times of market turmoil where the VIX exhibits peaks, they even produce negative skewness for VIX options, which is at odds with empirical evidence, even if the VIX IVs tend to become flat during the crisis (see Figure 2, VIX IVs on 27/10/2008). The VIX negative skewness generated by the SVJ2 model is interestingly not visible in the static calibration presented in Section 6 and might be an issue that appears when trying to consistently estimate the model over a long period of time. Finally, in January 2009 the model implied kurtosis is more than three times equal to its market counterpart.

7.4 Information contained in the different data sources

In this part we address the two following questions: first, what information do levels contain on options and second, which information do S&P 500 and VIX options share?

Since the value of the VIX index is calculated using a portfolio of S&P 500 options, it is tempting to see that the VIX as a summary of the information contained in S&P 500 options. In fact, Figure 8 shows that the volatilities filtered are the same across estimation datasets before the crisis. This seems to contradict previous results (e.g., Johannes, Polson, and Stroud (2009)) stating that including options in the estimation dataset increases the estimated volatility level. We attribute this difference to the fact that our Dataset 1 is not only composed of returns (as in Johannes, Polson, and Stroud (2009)) but also of VIX index levels, which incorporates some information about the options

\[ \text{Insert Figure 13 here} \]

\[ \text{34} \]

\[ ^{15} \text{However, we keep in mind that this result might be driven to some extent by the small sample of VIX options at our disposal, when the non-parametric formula of Bakshi, Kapadia, and Madan (2003) that we use ideally requires a continuum of options.} \]
market. In fact, our results suggest that in the calm periods, Dataset 1 is sufficient to recover the trajectory of the volatility. Following such reasoning, it is interesting to see to which extent the model estimated using Dataset 1 is able to reproduce options’ prices. We obtain RMSEs and RMSREs that are respectively more than twice those obtained using Dataset 4 for both S&P 500 and VIX options (even in the calmer first period). Our results indicate that even though estimating the model to returns and VIX index time series seems to be sufficient to filter volatility in "good times", it is definitely not sufficient to estimate appropriately model parameters. This leads to strong mispricing of options in both the SPX and VIX markets, even in the in-sample period. In contrast, in times of crisis, we see that the trajectories of volatility can differ significantly depending on the estimation dataset. In turn, Figure 15 shows that the central tendency process is consistently estimated across datasets during quiet periods, however in times of market turmoil the different datasets do not contain similar information on its trajectory. Including options into the estimation datasets leads to a strong increase in $m$ at the end of 2008 after the Lehman Brothers bankruptcy as well as during the Greek sovereign debt crisis, which is not detected when calibrating the model to Dataset 1. We conclude that estimating the model to underlying index values is not sufficient to reproduce option prices in either market and to extract the dynamics of the central tendency process accurately.

![Insert Figure 15 here]

We have already mentioned that the trajectories of the volatility and the central tendency differ a lot in times of crisis depending on the estimation dataset. We additionally notice that VIX and S&P 500 options provide conflicting information on the jump component of $v$ and $m$. In particular, the S&P500 options’ dataset seems to infer jumps in the volatility for both the SVJ and SVJ2 models, with a large jump at the beginning of the crisis (fall 2008), and a jump in the central tendency when the Eurozone sovereign debt crisis emerged (May 2010). On the other hand, the VIX options dataset finds a small jump in volatility (around 2%) in 2007 but finds a large jump in the central tendency in the fall of 2008. Estimating the SVJ and SVJ2 models to Dataset 4 generally combines the jumps detected in the variance using the SPX and VIX datasets, but decreases the estimated size of jumps. In fact, Dataset 4 seem to rather attribute jumps to the variance process and not to the central tendency process.

Furthermore, we note that the RMSEs and RMSREs of VIX options using Dataset 2 are about twice as large as those using Dataset 3, see Tables 4 and 8. This consideration holds in- and out-of-sample.
Thus we conclude that S&P 500 options do not span the information contained in VIX options. Conversely, the RMSEs and RMSREs of S&P 500 options are overall much lower (reduced by a factor 2 and 4 approximately) using Dataset 2 than using Dataset 3. This is in particular due to deep OTM calls that are not well reproduced at all using Dataset 3, which indicates that VIX options contain less information on the tail of the returns' distribution than S&P 500 options. Concerning deep OTM puts on the S&P 500, it is striking to see that the estimation using Dataset 3 outperforms the one using Dataset 2, which indicates that VIX options provide valuable information on the left tail of the returns' distribution.

We however notice that there is a significant loss of quality in the fitting of VIX options when estimating the SVJ2 model to all data sources, which indicates that even the SVJ2 model is not flexible enough to reconcile the information contained in all data sources in a completely consistent way. This contrasts with the results obtained in the static calibrations performed in section 6.

### 7.5 Risk premia

We first analyze the signs of the risk premia defined in Section 4.2. The equity risk premium coefficient $\eta_Y$ is found to be positive throughout the models and datasets considered, which is in line with a positive diffusive equity risk premium. When options are part of the estimation dataset, it is consistently between 0.6 and 0.85. The mean price jump risk premium is found to be slightly negative when all data sources are reconciled, i.e., the mean jump size of returns is slightly more negative under $Q$ than under $P$, which is also what Pan (2002) finds. The volatility of price jump risk premium given by the full model is around 10%, indeed the volatility of jump sizes under $Q$ is around 10-15% while it is around 2-3% under $P$. The diffusive part of the volatility risk premium is found to be negative, its amplitude however depends on the model used. In particular, it has much smaller magnitude with the SVJ2 model than with the SVJ and SV2 models. Our results on the mean volatility jump risk premium are mitigated. It is indeed positive when using Dataset 3 but negative when using Dataset 4. The average jump size under $P$ increases when adding S&P 500 options while the reverse occurs under $Q$. However, the fact that no jump has been filtered using Dataset 3 suggests that mean volatility jump risk premium is negative. Finally, the diffusive volatility risk premium in $m$ is also found to be negative, i.e., the process means-reverts quicker under $P$ than under $Q$. We did not obtain conclusive results on the central tendency jump risk premium.
We now examine integrated risk premia and their term structure as implied by the different models, calibrated over the four datasets previously considered.

The annualized integrated equity risk premium $ERP_t$ at time $t$ is defined as:

$$ ERP_t = \frac{1}{T-t} \left[ E^p_t \left( \int_t^T dY_s \right) - E^Q_t \left( \int_t^T dY_s \right) \right]. $$

The annualized integrated variance risk premium $VRP_t$ can be decomposed into a continuous and a jump part as follows:

$$ VRP^c_t = \frac{1}{T-t} \left[ E^p_t \left( \int_t^T v_s ds \right) - E^Q_t \left( \int_t^T v_s ds \right) \right], $$

$$ VRP^j_t = \frac{1}{T-t} \left[ E^p_t \left( \sum_{t \leq s \leq T} (\Delta Y_s)^2 \right) - E^Q_t \left( \sum_{t \leq s \leq T} (\Delta Y_s)^2 \right) \right], $$

$$ VRP_t = VRP^c_t + VRP^j_t. $$

For a detailed discussion on risk premia we refer to Bollerslev and Todorov (2011). In our setup all risk premia are available in closed-form.

We obtain integrated equity risk premia that are positive and strongly increasing during the crisis in 2008, then coming back to a low level and suddenly increasing again end of 2010, following the VIX movements, see Figure 16. During these periods of market turmoil, shorter-maturity risk premia are larger than longer-maturity premia which is in line with Aït-Sahalia, Karaman, and Mancini (2012). Our results are consistent in sign and magnitude across estimation datasets. Furthermore we find that the equity risk premium is primarily determined by its continuous part. Indeed, the jump part of the daily risk premium is usually smaller than 0.1 while the continuous part reaches about 0.7.

Finally, we obtain integrated variance risk premia which are slightly negative, with a magnitude which increases in 2008. Consistently with the literature shorter-maturity risk premia are close to zero than longer-maturity premia. The one year risk premium reaches -0.05 during the volatility peak and is mainly determined by its jump component when using the SVJ and SVJ2 models.

[Insert Figure 16 here]
8 Conclusion

In this paper we estimate a general affine model with jumps using a time series of S&P 500 and VIX levels as well as option prices on both indices. To extract as much information about extreme events as possible, we use S&P 500 and VIX options with a unique wide range of moneynesses. We depart from most of the literature and estimate the historical $P$-parameters and the risk-neutral $Q$-parameters jointly, in a single step. This estimation puts less restrictions on the parameters and allows us to obtain results on risk premia which are more data-driven. We show that although the standard SVJ model performs well at representing the smiles of volatility for both markets on a given date, its dynamics is not sufficiently flexible to accommodate for the dynamical properties embedded in the time series of option prices. We argue that a model with a stochastic central tendency and jumps in the returns and in each volatility factor brings significant improvements and allows to reach smaller pricing errors, both in and out-of-sample. We analyze the filtered trajectories of the latent factors using different estimation datasets and sub-models. We show that the variance process exhibits large and fast variations capturing the short-term movements of the volatility while the stochastic central tendency exhibits more persistence and hence reflects long-term expectations of investors. We provide an extensive analysis of which features of the SVJ2 are needed to represent different datasets. In particular, likelihood criteria as well as statistical tests conclude that the whole flexibility of the model is needed to jointly represent the index levels and the derivatives’ prices on both markets. Indeed, adding a stochastic central tendency helps to better represent the tails of the returns’ distribution as well as the term structure of S&P 500 and VIX option prices. Furthermore, jumps enable to better reflect the tail of the variance distribution. We highlight the limitations of the models considered, in particular we show that they are not able to fully reproduce the skewness and kurtosis of the underlying instruments in times of market turmoil. We investigate the information contained in the underlying levels and in the options on both markets. One the one hand, we find that the VIX index does not provide an accurate representation of the information contained in S&P 500 options, and on the other hand that the information contained in S&P 500 derivatives does not span the information contained in VIX derivatives and vice-versa. It is therefore crucial to include underlyings as well as derivatives on both markets to estimate a model and account for the cross section of instruments. We finally provide a discussion on the risk premia embedded in the model. We find that all the datasets considered provide consistent information on the equity risk premium

38
and that it is mainly determined by its continuous component. We obtain variance risk premia which are slightly negative and on the contrary mainly affected by their jump part.
References


Computational Finance, 2(4), 1–18.


Fang, F., and C. Oosterlee, 2008, “A novel pricing method for European options based on Fourier-
cosine series expansions,” working paper, MPRA Paper 9319, University Library of Munich, Germany.


Appendix

A Affine dependence of the VIX\(^2\) on \(v\) and \(m\)

We provide the expression of the coefficients \(\alpha_{VIX}^2\), \(\beta_{VIX}^2\) and \(\gamma_{VIX}^2\) in Proposition 4.1. We can write

\[
\alpha_{VIX}^2 = (1 + 2\lambda_Y^v C) A
\]
\[
\beta_{VIX}^2 = (1 + 2\lambda_Y^v C) B + (2\lambda_Y^v C) \hat{A}
\]
\[
\gamma_{VIX}^2 = 2\lambda_Y^v C + (1 + 2\lambda_Y^v C) G + (2\lambda_Y^v C) \hat{B}.
\]

where \(C := \left( \theta_{QZ}(1,0,0) - \frac{\partial \theta_{QZ}(0,0,0)}{\partial \phi_Y}(0,0,0) - 1 \right)\) and coefficient \(A, B, \hat{A}, \hat{B}, G\) are defined in Table 11 of the online Appendix. Furthermore, \(A, B, \hat{A}, \hat{B}, G\) are functions of:

\[
a_m := \left( \frac{\partial \theta_{QZ}(0,0,0)}{\partial \phi_m} \lambda_Y^m \right),
\]
\[
c_m := \left( \kappa_m \theta_{QZ}^{(Q)}(0,0,0) \lambda_Y^m \right),
\]
\[
b_m := -\frac{c_m}{a_m}, \text{ when } a_m \neq 0,
\]
\[
a_v := \left( \frac{\partial \theta_{QZ}(0,0,0)}{\partial \phi_v} \lambda_Y^v \right),
\]
\[
b_v := b_m \left( \kappa_v + \frac{\partial \theta_{QZ}(0,0,0)}{\partial \phi_v} \lambda_Y^v \right) + \frac{\partial \theta_{QZ}(0,0,0)}{\partial \phi_v} \lambda_Y^v, \text{ when } a_m \neq 0,
\]
\[
h_v := \left( \kappa_v \theta_{QZ}^{(Q)}(0,0,0) \lambda_Y^v \right).
\]

These expressions remain valid with different specifications for the jump structure (dependent jumps, same Poisson processes, etc.) in the model (1) - (3).
B Coefficients for the Fourier Cosine Expansion

Here we give the expression for $U_{VIX}^{n}$, the Fourier cosine transform of VIX options' payoff. To ease notation, we drop the subscript $VIX$ for $a_{VIX}, b_{VIX}$ and define $\omega_n := \frac{n\pi}{b-a}$. $U_{VIX}^{n}$ is given by:

$$U_{VIX}^{n} = \int_a^b (\sqrt{x} - K)^+ \cos (\omega_n (x - a))\, dx$$

$$= \frac{2}{b-a} \text{Re} \left\{ e^{-i\omega_n a} \left[ \sqrt{b} \frac{e^{-i\omega_n b}}{i\omega_n} + \frac{\sqrt{\pi}}{2} \left( \frac{1}{(-i\omega_n)^{3/2}} \left( \text{erf}(\sqrt{1-i\omega_n b}) - \text{erf}(K\sqrt{1-i\omega_n K}) \right) \right) \right\}$$

for $n > 0$, \quad (39)

$$U_{0}^{VIX} = \frac{2}{b-a} \left[ \frac{2}{3} b^{3/2} - Kb + \frac{1}{3} K^3 \right]. \quad (40)$$

C Interpolation procedure for the S&P 500 options dataset

The interpolation method we have used to find a parametric fit for each slice of S&P 500 implied volatilities is developed in detail in Rebonato and Cardoso (2004). Here we give the main idea and the particular choice of parameters we have made so that the slices are well fitted.

The idea of Rebonato and Cardoso (2004) is to use a mixture of log-normal densities for the Futures price. Log-normal densities mixtures can fit various different shapes including multimodal densities arising from jumps in the Futures price process. The ability of the method to recover any type of density (or equivalently smile) is well documented in Rebonato and Cardoso (2004) and chapter 9 of Rebonato (2004).

The attractive feature of this parametric representation for the density of the Futures price is that the pricing of call/put options can be performed using a mixture of Black-Scholes price. Additionally, the no-arbitrage condition is simply a condition on the expectation for the mixture of the Futures price.

In practice, a mixture of 4 log-normal densities is enough to have a nearly perfect fit. We minimize the Euclidean distance between market and mixture option prices using the CMA-ES algorithm mentioned in section 5.1. We have checked that the resulting fits are satisfactory by computing...
different measures of the distance between the market and model implied volatility slices. For instance, the RMSE between implied volatilities of the parametric fit and the true implied volatilities is most of the time below 0.25%. Sometimes the RMSE is larger and goes up to 2%, however this is not due to the inability of the mixture of log-normals to fit an implied volatility slice but due to the shape of the data. This is best explained by looking at a typical fit as displayed on Figure 6. We can see that the fit is nearly perfect, however the RMSE is not so close to zero because the input data is too rough. This phenomenon is amplified if the data has a sawtooth pattern (potential arbitrage) although the fit is very good.

Finally, using the parametric fit, we can sample ”market option prices” for the desired strikes and moneynesses. We have chosen to re-sample option prices from each parametric slice uniformly in strike. We also choose not to resample options for which the strike is smaller than 40% or larger than 140% of the current Futures price. The reason is that there are usually only one or two options outside this interval of moneyness and we do not wish to re-sample options where the interpolation results could be driven by an outlier.

D Particle Filter

The filtering procedure consists in approximating the distribution \( p(L_{t_k} | y^{f_k}, \Theta, M) \) of unobservable (latent) factors \( L_{t_k} \) at every point in time \( t_k = k \Delta t \) given the available observations \( y^{f_k} = \{ y_{t_0}, ..., y_{t_k} \} \) and assuming that the model \( M \) is well specified and its parameters \( \Theta \) are known. This approximation is denoted by \( \hat{p}(L_{t_k} | y^{f_k}, \Theta, M) \) as in (37). In the remaining of this section, we drop the dependence on the parameter set \( \Theta \) and the model specification \( M \). The available market measurements are the S&P 500 daily returns, the VIX levels and the option prices on both indices. We refer to section 5.2 for notations. The algorithm can be decomposed into the following steps.

1. Initialization
   
   We simulate \( n_p \) initial particles for the latent variables \( \{ v_{t_0}^{(i)}, m_{t_0}^{(i)} \}_{i=1,...,n_p} \) which are compatible with the initial value of the VIX squared, i.e., given the specification (28).

The steps 2 to 5 described below are repeated for each time step \( t_k \) in the grid from \( k = 0 \) to \( k = M - 1 \).
2. First-stage resampling

At this point, we assume that we have $n_p$ particles (i.e., possible values of $m$ and $v$) at time $t_k$ given all observations $y^k$ up to $t_k$. At time $t_{k+1}$, there are new observations $y_{k+1}$. The goal of this step is to keep from the previous sample of particles $\{v_{t_k}^{(i)}, m_{t_k}^{(i)}\}_{1 \leq i \leq n_p}$ only those which are likely to generate the new observation $y_{k+1}$. For this purpose, we assign a weight (namely first-stage weights) to each particle which is proportional to the likelihood of new market observations $y_{k+1}$ given the value of the particle $L_{t_k}$ at time $t_k$. Intuitively, particles that are compatible with the new observations will be assigned larger weights than other particles.

To increase the speed of the first-stage resampling, we do not consider options as part of the observations $y_{t_k+1}$ (only in this step) and limit $y_{t_k+1}$ to the values of the indices.

The first-stage weight $\omega_{t_{k+1}}^{(i)}$ assigned to the $i^{th}$ particle $L_{t_k}^{(i)}$ at time $t_{k+1}$ is given by:

$$\omega_{t_{k+1}}^{(i)} = p(L_{t_k}^{(i)}|y_{t_{k+1}}) \propto p(y_{t_{k+1}}|L_{t_k}^{(i)})$$

where $p(y_{t_{k+1}}|L_{t_k}^{(i)})$ is the density of the observation vector $y_{t_{k+1}}$ given the values of the particle vector $L_{t_k}^{(i)}$. The importance weights $\{\omega_{t_{k+1}}^{(i)}\}_{1 \leq i \leq n_p}$ add up to 1 so that they define a proper probability distribution. Conditioning on the number of jumps in $Y$ (or $v$) and $m$ gives:

$$\omega_{t_{k+1}}^{(i)} \propto \sum_{j,l \in \mathbb{N}} p(y_{t_{k+1}}|L_{t_k}^{(i)}, \Delta N^v_{t_k}, \Delta N^m_{t_k}) \mathbb{P}(\Delta N^v_{t_k} = j, \Delta N^m_{t_k} = l).$$

Given that we use daily observations, we limit the possible number of jumps of the Poisson random variables $\Delta N^v_{t_k}, \Delta N^m_{t_k}$ to zero or one (Bernoulli approximation). We recall that the new observation is composed of the SPX returns and the VIX level $y_{t_{k+1}} = (\Delta Y_{t_{k+1}}, VIX^2_{t_{k+1}})$. Since the $VIX^2_{t_{k+1}}$ is the sum of normal distributions and up to two exponential distributions, there is no closed form for this bivariate density in the general case. To preserve tractability, we approximate the exponentially distributed jump sizes by a categorical distribution (generalization of a Bernoulli distribution) which has support a certain number of (the corresponding exponential distribution) quantiles. As a consequence, the weight $\omega_{t_{k+1}}^{(i)}$ is a sum of weighted bivariate normal densities.

To eliminate the particles $\{L_{t_k}^{(i)}\}_{1 \leq i \leq n_p}$ that are not likely to generate the new observations
we resample (with replacement) particles according to a stratified resampling scheme:

\[ z(i) \sim \text{StratRes}(n_p, \omega_1^{(1)}, ..., \omega_1^{(n_p)}). \]

This allows to create a new sample of \( n_p \) latent factors \( \{L_i^{(j)}\}_{1 \leq j \leq n_p} \) which are now equally likely. Indeed, particles representing \( m_{t_k} \) and \( v_{t_k} \) are shuffled into a new set of particles: \( \{m_{t_k}^{(j)}, v_{t_k}^{(j)}\}_{j=1..n_p} = \{m_{t_k}^{z(i)}, v_{t_k}^{z(i)}\}_{i=1..n_p} \). We resample the same number of particles although this is in principle not necessary.

The next step of the particle filter consists in propagating the latent factors according to their conditional density given the previous values \( L_t^{(i)} \) and the new observations \( y_{t_k+1} \):

\[ L_{t_k+1}^{(i)} \sim p(L_{t_k+1}^{(i)} | L_t^{(i)}, y_{t_k+1}). \]

Because the distribution \( p(L_{t_k+1}^{(i)} | L_t^{(i)}, y_{t_k+1}) \) is not known in closed-form, we use a proposal density \( q(L_{t_k+1}^{(i)} | L_t^{(i)}, y_{t_k+1}) \). Propagating \( v \) and \( m \) requires preliminary knowledge on the jump components so we first focus on the jumps.

3. Generating the jumps

We calculate the joint probability of jumps in \( Y \) (or \( v \)) and \( m \) between \( t_k \) and \( t_{k+1} \) using:

\[ P(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{Nm} | y_{t_k+1}) \propto p(y_{t_k+1} | \Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{Nm}) P(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{Nm}). \tag{41} \]

Conditionally on the jump sizes in \( v \) and \( m \), the first part of the right hand-side has already been calculated in the first-stage weights. Using Bayes’ rule, we get an approximation for \( P(\Delta N_{t_k}^{Yv}, \Delta N_{t_k}^{Nm} | y_{t_k+1}) \). However, to have a better chance at detecting extreme events, we force the probability of jumps to be at least 10% for all processes and simulate from the resulting distribution function the jumps.

We infer the jump size in the returns following Johannes, Polson, and Stroud (2009):

\[ y_{t_k+1} \]

\[ ^{17} \text{We checked that using a multinomial or stratified resampling scheme give the same results.} \]
\( Z_{tk}^{Y(i)} | \Delta N_{tk}^{Yv}, y_{tk+1} \) is normally distributed \( \mathcal{N}(\mu_{J}^{Y(i)}, \sigma_{J}^{Y(i)}) \) where \( \mu_{J}^{Y(i)} \) and \( \sigma_{J}^{Y(i)} \) are given by

\[
\begin{split}
(\sigma_{J}^{Y(i)})^2 = & \left( \frac{1}{\Delta N_{tk}^{Yv(i)}} + \frac{1}{\hat{v}_{tk+1}^{(i)}} \right)^{-1} \\
\mu_{J}^{Y(i)} = & \left( \sigma_{J}^{Y(i)} \right)^2 \frac{Y_{tk+1} - \hat{v}_{tk+1}^{(i)}}{\hat{v}_{tk+1}^{(i)}} + \frac{(\sigma_{J}^{Y(i)})^2}{\sigma_{J}^2} \mu_Y
\end{split}
\]

where \( \hat{v}_{tk+1}^{(i)} \) is an estimate of \( v_{tk+1} \) given the information we have up to time \( t_k \) and particle \( i \); we use \( \hat{v}_{tk+1}^{(i)} = \mathbb{E}[v_{tk+1}|v_{tk}] \) and

\[
\hat{\mu}_{Y}^{(i)} = - \left( \lambda_Y (\theta Q Z(1,0,0) - 1) + \frac{1}{2} v_{tk}^{(i)} - \Delta N_{tk}^{Yv(i)} \right) \Delta t.
\]

Finally, we simulate jump sizes for \( v \) and \( m \) according to their exponential law.

4. Propagating the volatility and central tendency

The latent factors \( v \) and \( m \) are propagated following a Milstein discretization scheme of the SDE. We use the full truncation method to prevent them from taking negative values.

5. Computing the filtering density

At this point, the newly generated particles \( \{L_{tk+1}^{(i)}\}_{1 \leq i \leq n_p} \) are a sample of \( p(L_{tk+1}|y^{tk+1}) \).

We now calculate the second-stage weights \( \{\pi_{tk+1}^{(i)}\}_{1 \leq i \leq n_p} \) which approximate the probabilities \( p(L_{tk+1}^{(i)}|y^{tk+1}) \) and give an approximation for the filtering density at time \( t_{k+1} \). These weights are proportional to the likelihood of observations at time \( t_{k+1} \) given the propagated particles \( L_{tk+1}^{(i)} \), with a correction related to the proposal density:

\[
\pi_{tk+1}^{(i)} \propto \frac{p(L_{tk+1}^{(i)}|L_{tk}^{(i)}) p(y_{tk+1}|L_{tk+1}^{(i)})}{\omega_{tk+1}^{(i)} q(L_{tk+1}^{(i)}|L_{tk}^{(i)}, y_{tk+1})}.
\]

The posterior distribution of state variables is approximated by:

\[
\hat{p}(L_{tk+1}|y^{tk+1}) = \sum_{i=1}^{n_p} \pi_{tk+1}^{(i)} \delta(L_{tk+1} - L_{tk+1}^{(i)})
\]
We choose the most-likely value of a given factor by taking the expectation of the estimated filtering density, e.g., $\hat{v}_{t_{k+1}} = \mathbb{E}[v_{t_{k+1}}^{(i)}]$.

The algorithm described above extracts latent factors if one assumes that the model parameters are known. Pitt (2002) builds on Gordon, Salmond, and Smith (1993) to show that the parameters can be estimated using the Maximum Likelihood Importance Sampling Criterion, defined as the product over time of the averages of the second-stage weights. The likelihood of observations given the values of particles is then estimated by the average of the second-stage weights over particles:

$$p(y^{t_{n}}|\Theta, M) = \prod_{k=1}^{M} \hat{p}(y_{t_{k}}|y^{f_{k-1}}, \Theta, M) \hat{p}(y_{t_{0}}|\Theta, M)$$

where

$$\hat{p}(y_{t_{k}}|y^{f_{k-1}}, \Theta, M) = \frac{1}{n_{p}} \sum_{i=1}^{n_{p}} \pi_{t_{k}}^{(i)}.$$
### Tables and Figures

#### E.1 Tables

**Table 1: Descriptive Statistics for S&P 500 Futures log-returns and VIX levels**

<table>
<thead>
<tr>
<th></th>
<th>March 2006 - November 2008</th>
<th>December 2008 - October 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard dev.</td>
</tr>
<tr>
<td>Log-returns SPX</td>
<td>-0.0007</td>
<td>0.0159</td>
</tr>
<tr>
<td>VIX</td>
<td>0.2044</td>
<td>0.1208</td>
</tr>
</tbody>
</table>

**Table 2: Quantiles of implied volatility values for S&P 500 and VIX options. The quantiles are calculated using all maturities and moneyness available from March 2006 to October 2010.**

<table>
<thead>
<tr>
<th>Quantiles</th>
<th>SPX IVs</th>
<th>VIX IVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>16.93%</td>
<td>62.95%</td>
</tr>
<tr>
<td>50%</td>
<td>23.16%</td>
<td>75.03%</td>
</tr>
<tr>
<td>75%</td>
<td>32.77%</td>
<td>91.68%</td>
</tr>
</tbody>
</table>

**Table 3: Root Mean Squared Error in implied volatilities when calibrating the SVJ and the SVJ2 models to S&P 500 and VIX options on four different dates. On each date, the distance (23) is minimized using a global optimizer over the model parameters under $Q$. On each date, the data is composed of all S&P 500 options implied volatilities and all VIX implied volatilities together. Here we report the distances $RMSE_{SPX}$ and $RMSE_{VIX}$ but the minimization is run over $RMSE = \frac{1}{2}(RMSE_{SPX} + RMSE_{VIX}).$**

<table>
<thead>
<tr>
<th>RMSE (%)</th>
<th>20070711</th>
<th>20081022</th>
<th>20090610</th>
<th>20100505</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPX</td>
<td>VIX</td>
<td>SPX</td>
<td>VIX</td>
</tr>
<tr>
<td>SVJ2</td>
<td>0.852</td>
<td>3.885</td>
<td>2.427</td>
<td>12.757</td>
</tr>
</tbody>
</table>
Table 4: Root Mean Square Pricing Errors (on SPX and VIX options) for the SVJ2, the SVJ and the SV2 models depending on the dataset these models were estimated to, and conditional on the moneyness and maturity of options. The last column with heading # describes the number of options in each category given in the first column. The moneyness is defined here as the ratio between the strike of the option and the corresponding maturity futures’ price. Deep OTM puts on the SPX refer to put options with moneyness smaller than 0.7, OTM puts on the SPX have a moneyness between 0.7 and 0.95, close to ATM options have a moneyness between 0.95 and 1.05, OTM calls between 1.05 and 1.2 and deep OTM calls above 1.2. Deep OTM calls on the VIX have moneyness above 1.3, OTM calls between 1.1 and 1.3, close to ATM options between 0.9 and 1.1 and ITM calls below 0.9. Short maturity options refer to options with time-to-maturity smaller than two months. Medium maturity options on the SPX denote options with time-to-maturity between two and six months and long maturity options above six months. Medium maturity options on the VIX denote options with time-to-maturity larger than two months.

<table>
<thead>
<tr>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVJ2</td>
<td>SVJ</td>
<td>SV2</td>
<td>SVJ2</td>
</tr>
<tr>
<td>Close to ATM options</td>
<td>13.604</td>
<td>7.731</td>
<td>10.880</td>
</tr>
<tr>
<td>Short mat options</td>
<td>2.619</td>
<td>2.073</td>
<td>2.457</td>
</tr>
<tr>
<td>Medium mat options</td>
<td>7.856</td>
<td>5.384</td>
<td>6.428</td>
</tr>
</tbody>
</table>

| Short mat options | 2.732    | 2.219     | 2.609     | 1.545     | 1.762     | 1.579     | 3.009     | 3.765     | 3.442     |

| Overall   | 0.021     | 0.026     | 0.022     | 0.017     | 0.024     | 0.019     | 0.010     | 0.015     | 0.013     |
| Deep OTM puts | 0.008    | 0.007     | 0.007     | 0.006     | 0.006     | 0.006     | 0.004     | 0.004     | 0.005     |
| OTM puts  | 0.014     | 0.013     | 0.012     | 0.011     | 0.012     | 0.011     | 0.004     | 0.007     | 0.007     |
| Close to ATM options | 0.021   | 0.020     | 0.018     | 0.015     | 0.019     | 0.015     | 0.005     | 0.013     | 0.009     |
| OTM calls | 0.032     | 0.046     | 0.039     | 0.029     | 0.042     | 0.034     | 0.019     | 0.027     | 0.023     |
| Short mat options | 0.025    | 0.022     | 0.020     | 0.013     | 0.020     | 0.016     | 0.011     | 0.015     | 0.012     |
| Medium mat options | 0.025    | 0.029     | 0.024     | 0.020     | 0.027     | 0.022     | 0.009     | 0.015     | 0.013     |

| Overall   | 0.025     | 0.039     | 0.032     | 0.021     | 0.038     | 0.026     | 0.014     | 0.024     | 0.017     |
| Deep OTM puts | 0.007    | 0.009     | 0.009     | 0.009     | 0.009     | 0.009     | 0.006     | 0.006     | 0.007     |
| OTM puts  | 0.014     | 0.020     | 0.018     | 0.016     | 0.019     | 0.016     | 0.006     | 0.014     | 0.011     |
| Close to ATM options | 0.025   | 0.035     | 0.029     | 0.022     | 0.034     | 0.024     | 0.009     | 0.025     | 0.016     |
| OTM calls | 0.046     | 0.076     | 0.059     | 0.036     | 0.073     | 0.048     | 0.027     | 0.044     | 0.032     |
| Short mat options | 0.022    | 0.032     | 0.028     | 0.017     | 0.031     | 0.022     | 0.010     | 0.017     | 0.014     |
| Medium mat options | 0.027    | 0.044     | 0.034     | 0.024     | 0.043     | 0.029     | 0.016     | 0.029     | 0.019     |
Table 5: Estimated parameters from the particle filter for the different models and datasets.

<table>
<thead>
<tr>
<th></th>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVJ2</td>
<td>SVJ</td>
<td>SV2</td>
<td>SVJ2</td>
</tr>
<tr>
<td>( \lambda^Y_p )</td>
<td>0.394</td>
<td>0.030</td>
<td>0</td>
<td>2.722</td>
</tr>
<tr>
<td>( \lambda^Y_v )</td>
<td>1.650</td>
<td>0.007</td>
<td>0</td>
<td>0.079</td>
</tr>
<tr>
<td>( \lambda^Y_m )</td>
<td>2.438</td>
<td>0</td>
<td>0</td>
<td>0.106</td>
</tr>
<tr>
<td>( \lambda^V_p )</td>
<td>0.134</td>
<td>0</td>
<td>0</td>
<td>0.080</td>
</tr>
<tr>
<td>( \lambda^V_v )</td>
<td>0.070</td>
<td>0</td>
<td>0</td>
<td>0.644</td>
</tr>
<tr>
<td>( \sigma_m )</td>
<td>0.150</td>
<td>0</td>
<td>0.147</td>
<td>0.113</td>
</tr>
<tr>
<td>( \sigma_v )</td>
<td>0.715</td>
<td>0.718</td>
<td>0.717</td>
<td>0.951</td>
</tr>
<tr>
<td>( \rho_{Y,V} )</td>
<td>-0.550</td>
<td>-0.625</td>
<td>-0.578</td>
<td>-0.910</td>
</tr>
<tr>
<td>( m_0 )</td>
<td>0.027</td>
<td>0.020</td>
<td>0.023</td>
<td>0.028</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa^F_v )</td>
<td>0.060</td>
<td>0</td>
<td>0.057</td>
<td>0.212</td>
</tr>
<tr>
<td>( \theta^F_p )</td>
<td>0.501</td>
<td>0</td>
<td>0.504</td>
<td>0.036</td>
</tr>
<tr>
<td>( \nu^F_p )</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
<td>0.127</td>
</tr>
<tr>
<td>( \nu^F_v )</td>
<td>0.112</td>
<td>0.155</td>
<td>0</td>
<td>0.265</td>
</tr>
<tr>
<td>( \mu^F_p )</td>
<td>-0.060</td>
<td>0.150</td>
<td>0</td>
<td>-0.209</td>
</tr>
<tr>
<td>( \sigma^F_p )</td>
<td>0.041</td>
<td>0.007</td>
<td>0</td>
<td>0.475</td>
</tr>
<tr>
<td>( \eta^F )</td>
<td>0.721</td>
<td>0.000</td>
<td>0.733</td>
<td>0.150</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa^Q_v )</td>
<td>0.111</td>
<td>0</td>
<td>0.108</td>
<td>0.383</td>
</tr>
<tr>
<td>( \theta^Q_p )</td>
<td>0.272</td>
<td>0</td>
<td>0.266</td>
<td>0.020</td>
</tr>
<tr>
<td>( \nu^Q_p )</td>
<td>0.114</td>
<td>0</td>
<td>0</td>
<td>0.020</td>
</tr>
<tr>
<td>( \nu^Q_v )</td>
<td>0.157</td>
<td>0.763</td>
<td>0</td>
<td>0.010</td>
</tr>
<tr>
<td>( \mu^Q_p )</td>
<td>0.040</td>
<td>-0.120</td>
<td>0</td>
<td>0.020</td>
</tr>
<tr>
<td>( \sigma^Q_p )</td>
<td>0.033</td>
<td>0.006</td>
<td>0</td>
<td>0.006</td>
</tr>
</tbody>
</table>

|                       |           |           |           |           |
| Log-likelihood       | 10626     | 10572     | 10520     | 9785      | 4539      | 8238      | 9893      | 8220      | 9123      | 5133      | -1599     | 2239      |
| AIC                  | -21206    | -21116    | -21020    | -19518    | -9044     | -16450    | -19730    | -16402    | -18216    | -10204    | -3242     | -4442     |
| BIC                  | -21103    | -21053    | -20975    | -19402    | -8968     | -16392    | -19605    | -16317    | -18149    | -10065    | -3341     | -4361     |
Table 6: Statistics on in-sample and out-of-sample errors on VIX levels, for the different models and datasets.

<table>
<thead>
<tr>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVJ2</td>
<td>SV J</td>
<td>SV2</td>
<td>SVJ2</td>
</tr>
<tr>
<td>In-sample fitting levels</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean error VIX</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>RMSE VIX</td>
<td>0.023</td>
<td>0.023</td>
<td>0.022</td>
</tr>
<tr>
<td>Out-of-sample fitting levels</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean error VIX</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>RMSE VIX</td>
<td>0.022</td>
<td>0.022</td>
<td>0.021</td>
</tr>
</tbody>
</table>

Table 7: Diebold-Mariano test values for in-sample and out-sample errors on VIX levels, for different models calibrated to log-returns and VIX levels. Two loss functions are considered: the absolute and quadratic error. Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

<table>
<thead>
<tr>
<th>SVJ</th>
<th>SV2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abs loss function</td>
<td>Quad loss function</td>
</tr>
<tr>
<td>In-sample</td>
<td>-0.278</td>
</tr>
<tr>
<td>Out-of-sample</td>
<td>-0.243</td>
</tr>
</tbody>
</table>
Table 8: Root Mean Square Relative Pricing Errors (on SPX and VIX options) for the SVJ2, the SVJ and the SV2 models depending on the dataset these models were estimated on, and conditional on the moneyness and maturity of options. The last column with heading # describes the number of options in each category given in the first column. The moneyness is defined here as the ratio between the strike of the option and the corresponding maturity futures’ price. Deep OTM puts on the SPX refer to put options with moneyness smaller than 0.7, OTM puts on the SPX have a moneyness between 0.7 and 0.95, close to ATM options have a moneyness between 0.95 and 1.05, OTM calls between 1.05 and 1.2 and deep OTM calls above 1.2. Deep OTM calls on the VIX have moneyness above 1.3, OTM calls between 1.1 and 1.3, close to ATM options between 0.9 and 1.1 and ITM calls below 0.9. Short maturity options refer to options with time-maturity smaller than two months. Medium maturity options on the SPX denote options with time-to-maturity between two and six months and long maturity options above six months. Medium maturity options on the VIX denote options with time-to-maturity larger than two months.

<table>
<thead>
<tr>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fitting SPX</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>2.946</td>
<td>0.850</td>
<td>1.849</td>
</tr>
<tr>
<td>Deep OTM puts</td>
<td>0.738</td>
<td>0.625</td>
<td>0.573</td>
</tr>
<tr>
<td>OTM puts</td>
<td>0.501</td>
<td>0.398</td>
<td>0.449</td>
</tr>
<tr>
<td>Close to ATM options</td>
<td>0.675</td>
<td>0.367</td>
<td>0.484</td>
</tr>
<tr>
<td>OTM calls</td>
<td>5.221</td>
<td>1.779</td>
<td>3.246</td>
</tr>
<tr>
<td>Short mat options</td>
<td>1.785</td>
<td>0.979</td>
<td>1.151</td>
</tr>
<tr>
<td>Medium mat options</td>
<td>2.798</td>
<td>0.797</td>
<td>1.725</td>
</tr>
<tr>
<td>Long mat options</td>
<td>3.922</td>
<td>0.767</td>
<td>2.473</td>
</tr>
<tr>
<td><strong>Fitting SPX options in the in-sample period</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>2.337</td>
<td>0.751</td>
<td>1.631</td>
</tr>
<tr>
<td>Deep OTM puts</td>
<td>0.729</td>
<td>0.898</td>
<td>0.812</td>
</tr>
<tr>
<td>OTM puts</td>
<td>0.455</td>
<td>0.390</td>
<td>0.482</td>
</tr>
<tr>
<td>Close to ATM options</td>
<td>0.307</td>
<td>0.379</td>
<td>0.298</td>
</tr>
<tr>
<td>OTM calls</td>
<td>2.744</td>
<td>1.006</td>
<td>1.988</td>
</tr>
<tr>
<td>Short mat options</td>
<td>1.427</td>
<td>0.992</td>
<td>1.096</td>
</tr>
<tr>
<td>Medium mat options</td>
<td>2.447</td>
<td>0.609</td>
<td>1.633</td>
</tr>
<tr>
<td>Long mat options</td>
<td>2.783</td>
<td>0.674</td>
<td>1.982</td>
</tr>
<tr>
<td><strong>Fitting VIX options in the in-sample period</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>1.088</td>
<td>0.738</td>
<td>0.795</td>
</tr>
<tr>
<td>Deep OTM calls</td>
<td>1.309</td>
<td>0.764</td>
<td>0.835</td>
</tr>
<tr>
<td>OTM calls</td>
<td>1.218</td>
<td>0.811</td>
<td>0.898</td>
</tr>
<tr>
<td>Close to ATM options</td>
<td>1.054</td>
<td>0.751</td>
<td>0.867</td>
</tr>
<tr>
<td>ITM calls</td>
<td>0.629</td>
<td>0.609</td>
<td>0.514</td>
</tr>
<tr>
<td>Short mat options</td>
<td>1.101</td>
<td>0.761</td>
<td>0.909</td>
</tr>
<tr>
<td>Medium mat options</td>
<td>1.075</td>
<td>0.717</td>
<td>0.677</td>
</tr>
<tr>
<td><strong>Fitting VIX options in the out-of-sample period</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.650</td>
<td>0.894</td>
<td>0.793</td>
</tr>
<tr>
<td>Deep OTM calls</td>
<td>0.714</td>
<td>0.916</td>
<td>0.946</td>
</tr>
<tr>
<td>OTM calls</td>
<td>0.652</td>
<td>0.915</td>
<td>0.756</td>
</tr>
<tr>
<td>Close to ATM options</td>
<td>0.646</td>
<td>0.873</td>
<td>0.675</td>
</tr>
<tr>
<td>ITM calls</td>
<td>0.505</td>
<td>0.855</td>
<td>0.588</td>
</tr>
<tr>
<td>Short mat options</td>
<td>0.666</td>
<td>0.846</td>
<td>0.798</td>
</tr>
<tr>
<td>Medium mat options</td>
<td>0.637</td>
<td>0.930</td>
<td>0.789</td>
</tr>
</tbody>
</table>
Table 9: Diebold-Mariano test values for in-sample and out-sample errors on VIX option prices, for the different models and estimation datasets. Two loss functions are considered namely the average Mean Square Error (MSE) and the average Mean Square Relative Error (MSRE). Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

<table>
<thead>
<tr>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVJ</td>
<td>SV2</td>
</tr>
<tr>
<td>MSE</td>
<td>MSRE</td>
</tr>
<tr>
<td>In-sample</td>
<td>0.000</td>
</tr>
<tr>
<td>Out-of-sample</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 10: Diebold-Mariano test values for in-sample and out-sample errors on S&P 500 option prices, for different models and estimation datasets. Two loss functions are considered namely the average Mean Square Error (MSE) and the average Mean Square Relative Error (MSRE). Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

<table>
<thead>
<tr>
<th>Dataset 2</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVJ</td>
<td>SV2</td>
</tr>
<tr>
<td>MSE</td>
<td>MSRE</td>
</tr>
<tr>
<td>In-sample</td>
<td>28.600</td>
</tr>
<tr>
<td>Out-of-sample</td>
<td>60.574</td>
</tr>
</tbody>
</table>
E.2 Figures

Figure 1: Joint evolution of the VIX (dashed curve) and SPX index (solid curve) values from February 2006 to June 2010. The left y-axis corresponds to the VIX values (in percentage) and the right y-axis to the S&P 500 index values.
Figure 2: SPX and VIX implied volatility skews on four different dates as a function of log-moneyness ($\log K/F$). For each market, the scale is the same across days. The maturities $T$ are quoted in years.
Figure 3: One month expected returns of the S&P 500 Futures (Panel A) and VIX index (Panel B) implied by S&P 500 and VIX options with maturity one month from March 1st, 2006 to October 29th, 2010. We use the method described in Bakshi, Kapadia, and Madan (2003). Returns are expressed in percentage units.
Figure 4: Market and Heston model implied volatilities for VIX options (four maturities) on October 20th, 2010 plotted with respect to forward log-moneyness ($\log K/F(T)$). The market (resp. model) implied volatilities are represented by the crosses (resp. the solid line). These fits are obtained by minimizing relative errors between market implied volatilities and the Heston model implied volatility.
Figure 5: Comparative fit of 4 different maturities of VIX options for the SVJ2 (panels A to D) and the SVJ models (panels E to H) on July 11th, 2007. The crosses are the market implied volatilities and the curve represents the model volatilities. The implied volatilities are plotted as a function of forward log-moneyness (log $K/F(T)$).
Figure 6: Typical interpolation of market implied volatilities (circles) using a mixture of log-normal densities for the density of Futures prices. The implied volatilities are plotted as a function of forward log-moneyness ($\log K/F(T)$).

![Log-normal mixture fit](image)

Figure 7: Filtered trajectories of the latent factor $\sqrt{v}$ when estimating the SVJ2 (solid line), the SVJ (dashed line) and the SV2 (dashed dotted line) models over the S&P 500 log-returns, the VIX index values, VIX and S&P 500 option prices from March 2006 to November 2008 (685 days). The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

![Filtered trajectories](image)
Figure 8: Difference between the filtered volatilities $\sqrt{\nu_t(Dataset \ 4)} - \sqrt{\nu_t(Dataset \ j)}$ when estimating the SVJ2 model to different datasets $j = 1, 2, 3$; Dataset 4 being the benchmark. Panel A displays the difference with Dataset 1, Panel B with Dataset 2 and Panel C with Dataset 3. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.
Figure 9: Filtered jump sizes in the variance process $v$ when estimating the SVJ2 (solid line) and the SVJ (dashed line) models over the different datasets from March 2006 to November 2008 (685 days). We consider that there is a jump when the filtered probability of jump is larger than 50%. Panel A corresponds to Dataset 1 which comprises the underlying forward returns on the S&P 500 and the VIX levels. Panel B corresponds to Dataset 2 which consists of the underlying index levels plus S&P 500 options. Panel C corresponds to Dataset 3 which comprises the underlying index levels plus VIX options. Finally Panel D corresponds to Dataset 4 which gathers all data sources considered. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.
Figure 10: Filtered jump sizes in the process $m$ when estimating the SVJ2 model over the datasets 1 to 4 from March 2006 to November 2008 (685 days). We consider that there is a jump when the filtered probability of jump is larger than 50%. Each panel shows the jumps detected when estimating the SVJ2 model to a different dataset (Panel A: Dataset 1, ..., Panel D: Dataset 4). The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.
Figure 11: Filtered trajectories of the latent factor $m$ when estimating the SVJ2 (solid line), SVJ (horizontal dashed line) and the SV2 (dashed dotted line) model over the different datasets from March 2006 to November 2008 (685 days). Each panel corresponds to a different dataset. Dataset 1 is composed of SPX and VIX indices, Dataset 2 of both indices and SPX options, Dataset 3 of both indices and VIX options, Dataset 4 of both indices and both SPX/VIX options. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.
Figure 12: Fitting of the VIX index values for the SVJ2 model when the model is calibrated to log-returns and VIX levels (Dataset 1) from March 2006 to November 2008 (685 days). The crosses represent market data, the line the filtered values. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

Figure 13: 1 month risk-neutral skewness and kurtosis of the distribution of returns implied by 1 month SPX options prices when estimating the SVJ2 (solid line), the SVJ (dashed line) and the SV2 (dashed dotted line) models over Dataset 4 (indices as well as SPX and VIX options) from March 2006 to November 2008 (685 days). We use the method described in Bakshi, Kapadia, and Madan (2003). The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.
Figure 14: 1 month risk-neutral skewness and kurtosis of the distribution of the VIX implied by 1 month VIX options prices when estimating the SVJ2 (solid line), the SVJ (dashed line) and the SV2 (dashed dotted line) models over Dataset 4 (indices as well as SPX and VIX options) from March 2006 to November 2008 (685 days). We use the method described in Bakshi, Kapadia, and Madan (2003). The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.
Figure 15: Filtered trajectories of the latent factor $m$ when estimating the SVJ2 model over Dataset 1 (solid line), Dataset 2 (dashed line), Dataset 3 (dashed-dotted line) and Dataset 4 (dotted line) from March 2006 to November 2008 (685 days). Dataset 1 is composed of SPX and VIX indices, Dataset 2 of both indices and SPX options, Dataset 3 of both indices and VIX options, Dataset 4 of both indices and both SPX/VIX options. The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.

Figure 16: Equity and variance risk premia for different maturities when estimating the SVJ2 model using Dataset 4 from March 2006 to November 2008 (685 days). The shaded part of the graph represents the out-of-sample period, from 25 November 2008 until end of October 2010.
ONLINE APPENDIX TO
Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets

This appendix provides the results of technical derivations.
A Model specification under $P$

Under the historical measure $P$, the model is specified as follows:

$$
\begin{align*}
\mathrm{d}Y_t &= \left[ -\lambda Y_P(t,v_t,m_t)(\theta_Z(1,0,0) - 1) - \frac{1}{2}v_t - \gamma_t \right] \mathrm{d}t + \sqrt{v_t} \mathrm{d}W^Y_P + \mathrm{d}J^Y_P \\
\mathrm{d}v_t &= \kappa^{(P)}_v \left( \frac{\kappa^{(Q)}_v}{\kappa^{(P)}_v} m_t - v_t \right) \mathrm{d}t + \sigma_v \sqrt{v_t} \mathrm{d}W^v_P + \mathrm{d}J^v_P \\
\mathrm{d}m_t &= \kappa^{(P)}_m (\theta^{(P)}_m - m_t) \mathrm{d}t + \sigma_m \sqrt{m_t} \mathrm{d}W^m_P + \mathrm{d}J^m_P 
\end{align*}
$$

(A.1) 

(A.2) 

(A.3)

with:

$$
\begin{align*}
\gamma_t &= \eta_Y v_t - \lambda Y (v_t,m_t)(\theta^{(P)}_Z(1,0,0) - \theta^{(Q)}_Z(1,0,0)) \\
\mathrm{d}W^Y_P &= \mathrm{d}W^Y_Q - \eta_Y \sqrt{v_t} \mathrm{d}t \\
\mathrm{d}W^v_P &= \mathrm{d}W^v_Q + \sqrt{v_t} \frac{\kappa^{(P)}_v - \kappa^{(Q)}_v}{\sigma_v} \mathrm{d}t \\
\mathrm{d}W^m_P &= \mathrm{d}W^m_Q + \sqrt{m_t} \frac{\kappa^{(P)}_m - \kappa^{(Q)}_m}{\sigma_m} \mathrm{d}t \\
\theta^{(P)}_m &= \frac{\kappa^{(Q)}_m \theta^{(Q)}_m}{\kappa^{(P)}_m}.
\end{align*}
$$

B Characteristic functions

The characteristic function of the processes $Y$, $v$ and $m$ defined in the model (1) - (3) are exponential affine as stated in Proposition 4.2.

$$
\begin{align*}
\Psi_{VIX_t^2}(t,v,m;\omega) &= \mathbb{E}_t \left[ e^{\omega VIX^2_t} \right] = e^{\alpha(T-t)+\beta(T-t)v+\gamma(T-t)m}, \\
\Psi_{YT}(t,v,m;\omega) &= \mathbb{E}_t \left[ e^{\omega YT} \right] = e^{\alpha_Y(T-t)+\beta_Y(T-t)v+\gamma_Y(T-t)m+\delta_Y(T-t)m},
\end{align*}
$$

where $\omega \in \mathbb{C}$ is in each case chosen so that the integral converges.
The coefficients entering the definition of $Ψ_{VIX^2}$ satisfy the following ODEs:\textsuperscript{18}

\begin{align*}
- \alpha'(T-t) + \gamma(T-t)\kappa_m^{(Q)}\theta_m^{(Q)} + \lambda_0^Y v \left(\theta_Z^{(Q)}(0, \beta(T-t), 0) - 1\right) + \lambda_0^m \left(\theta_Z^{(Q)}(0, 0, \gamma(T-t)) - 1\right) & = 0 \\
- \beta'(T-t) - \beta(T-t)\kappa_m^{(Q)} + \frac{1}{2} \sigma_v^2 \beta^2(T-t) + \lambda_1^Y v \left(\theta_Z^{(Q)}(0, \beta(T-t), 0) - 1\right) & = 0 \\
- \gamma'(T-t) - \gamma(T-t)\kappa_m^{(Q)} + \frac{1}{2} \sigma_m^2 \gamma^2(T-t) + \kappa_v^{(Q)} \beta(T-t) + \lambda_2^Y v \left(\theta_Z^{(Q)}(0, \beta(T-t), 0) - 1\right) + \lambda_1^m \left(\theta_Z^{(Q)}(0, 0, \gamma(T-t)) - 1\right) & = 0
\end{align*}

$\forall t \in (0, T]$ with boundary conditions $\alpha(0) = 0$, $\beta(0) = \omega_1$ and $\gamma(0) = \omega_2$, where $\omega_1 := \omega\alpha_{VIX^2}$ and $\omega_2 := \omega\beta_{VIX^2}$ (the coefficients $\alpha_{VIX^2}$ and $\beta_{VIX^2}$ are defined in the Appendix A).

The coefficients entering the definition of $Ψ_{Y_t}$ satisfy the following ODEs are given by:

\begin{align*}
- \alpha_Y'(T-t) + \beta_Y(T-t)(-\lambda_0^Y v(\theta_Z^{(Q)}(1, 0, 0) - 1)) + \delta_Y(T-t)\kappa_m^{(Q)}\theta_m^{(Q)} \\
+ \lambda_0^Y v[\theta_Z^{(Q)}(\beta_Y(T-t), \gamma_Y(T-t), 0) - 1] + \lambda_0^m [\theta_Z^{(Q)}(0, 0, \delta_Y(T-t)) - 1] & = 0 \\
- \beta_Y'(T-t) & = 0 \\
- \gamma_Y'(T-t) - \beta_Y(T-t)\lambda_1^Y v(\theta_Z^{(Q)}(1, 0, 0) - 1) - \frac{1}{2} \beta_Y(T-t) - \gamma_Y(T-t)\kappa_v^{(Q)} + \frac{1}{2} \beta_Y(T-t)^2 \\
+ \frac{1}{2} \gamma_Y(T-t)^2 \sigma_v^2 + \beta_Y(T-t)\gamma_Y(T-t)\sigma_v \rho_{Y,v} + \lambda_1^Y v[\theta_Z^{(Q)}(\beta_Y(T-t), \gamma_Y(T-t), 0) - 1] & = 0 \\
- \delta_Y'(T-t) - \beta_Y(T-t)\lambda_2^Y v(\theta_Z^{(Q)}(1, 0, 0) - 1) + \gamma_Y(T-t)\kappa_v^{(Q)} - \delta_Y(T-t)\kappa_m^{(Q)} + \frac{1}{2} \delta_Y(T-t)^2 \sigma_m^2 \\
+ \lambda_2^Y v[\theta_Z^{(Q)}(\beta_Y(T-t), \gamma_Y(T-t), 0) - 1] + \lambda_1^m [\theta_Z^{(Q)}(0, 0, \delta_Y(T-t)) - 1] & = 0
\end{align*}

$\forall t \in (0, T]$ with boundary conditions $\alpha_Y(0) = 0$, $\beta_Y(0) = \omega$, $\gamma_Y(0) = 0$ and $\delta_Y(0) = 0$.

\textsuperscript{18}This relies on the fact that the Poisson processes driving the jumps in $v$ and in $m$ are independent.
C  Coefficients of the VIX\(^2\) formula

Table 11: Proposition 4.1 states that the VIX\(^2\) depends on the instantaneous variance and level of mean reversion in an affine way. Here we give the values of coefficients playing a role in this proposition (see Appendix A).

<table>
<thead>
<tr>
<th>Condition</th>
<th>(A)</th>
<th>(B)</th>
<th>(G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_m \neq 0) &amp; (a_v \neq 0) &amp; (a_m \neq a_m)</td>
<td>(\frac{1}{a_m VIX} (e^{a_m VIX} - 1))</td>
<td>(\frac{h_v}{a_m} \left( \frac{e^{a_m VIX} - 1}{a_m} - \left( \frac{e^{a_m VIX} - 1}{a_m} \right) \right) )</td>
<td>(b_m \left( \frac{e^{a_m VIX} - 1}{a_m} \right) - 1 - b_m B)</td>
</tr>
<tr>
<td>(a_m \neq 0) &amp; (a_v \neq 0) &amp; (a_v = a_m)</td>
<td>(\frac{1}{a_m VIX} (e^{a_v VIX} - 1))</td>
<td>(\frac{h_v}{a_m} \left( \frac{e^{a_v VIX} - 1}{a_v} - \left( \frac{e^{a_v VIX} - 1}{a_v} \right) \right) )</td>
<td>(\frac{h_v}{a_m} \left( \frac{e^{a_m VIX} - 1}{a_m} \right) - 1 - b_m B)</td>
</tr>
<tr>
<td>(a_m \neq 0) &amp; (a_v = 0)</td>
<td>(1)</td>
<td>(\frac{h_v}{a_m} \left( \frac{1}{a_m VIX} (e^{a_m VIX} - 1) - 1 \right) )</td>
<td>(\frac{1}{2} b_v VIX - b_m B)</td>
</tr>
<tr>
<td>(a_m = 0) &amp; (a_v \neq 0)</td>
<td>(\frac{1}{a_v VIX} (e^{a_v VIX} - 1))</td>
<td>(\frac{h_v}{a_v} \left( \frac{1}{a_v VIX} (e^{a_v VIX} - 1) - 1 \right) )</td>
<td>(c_m a_v \left( B - \frac{1}{2} h_v VIX \right) )</td>
</tr>
<tr>
<td>(a_m = 0) &amp; (a_v = 0)</td>
<td>(1)</td>
<td>(\frac{1}{2} VIX h_v)</td>
<td>(\frac{1}{2} VIX \left[ \frac{\partial h_v}{\partial v} (0, 0, 0) \lambda_0^v + c_m h_v \frac{VIX}{2} \right] )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>(A)</th>
<th>(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_m \neq 0)</td>
<td>(\frac{e^{a_m VIX} - 1}{a_m VIX})</td>
<td>(b_m (1 - A))</td>
</tr>
<tr>
<td>(a_m = 0)</td>
<td>(1)</td>
<td>(c_m VIX)</td>
</tr>
</tbody>
</table>