

Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets*

CHRIS BARDGETT[†] ELISE GOURIER[‡] MARKUS LEIPPOLD[§]

November 2016

Abstract

This paper shows that the VIX market contains information that is not already contained by the S&P 500 market on the variance of the S&P 500 returns. We estimate a flexible affine model based on a joint time series of underlying indexes and option prices on both markets. We find that including VIX option prices in the model estimation allows better identification of the parameters driving the risk-neutral conditional distributions and term structure of volatility, thereby enhancing the estimation of the variance risk premium. We gain new insights on the properties of the premium's term structure and show how they can be used to form trading signals. Finally, our premium has better predictive power than the usual model-free estimate and the higher-order moments of its term structure allow improving forecasts of S&P 500 returns.

KEYWORDS: S&P 500 and VIX joint modeling, volatility dynamics, particle filter, variance risk premium.

JEL CODES: G12, G13, C58.

*The authors thank Yacine Aït-Sahalia, Torben Andersen, Tim Bollerslev, Peter Christoffersen, Jérôme Detemple, Garland Durham, Damir Filipović, Andras Fulop, Michael Johannes, Lorian Mancini, Chay Ornthanalai, Chris Rogers, Olivier Scaillet, Ronnie Sircar, Josef Teichmann, Fabio Trojani, Anders Trolle and Alexandre Ziegler for helpful comments. Useful feedback was received from seminar participants at University of Zürich, ETH, University of Aarhus, HEC Montreal, Queen Mary University, Kent Business School, Toulouse School of Economics as well as at the Swiss Finance Institute Gerzensee workshop, the Bachelier Finance Conference, EFA 2013, the 2014 Conference of the Swiss Society for Financial Market Research, the 5th International Conference of the ERCIM WG on Computing & Statistics, and the 2015 French Econometrics Conference. Financial support from the Swiss Finance Institute (SFI), Bank Vontobel, the Swiss National Science Foundation and the National Center of Competence in Research "Financial Valuation and Risk Management" is gratefully acknowledged.

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

[‡]Queen Mary University of London and Center for Economic Policy Research (CEPR), Mile End Road, London E1 4NS, UK; Email: e.gourier@qmul.ac.uk.

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

1 Introduction

Introduced by the CBOE in 1993, the VIX index non-parametrically approximates the expected future realized volatility of the S&P 500 returns over the next 30 days. Options on the VIX started trading in 2006 and, as of today, represent a much larger market than VIX futures. By definition, the VIX index, VIX options, and S&P 500 options are directly linked to the S&P 500 index and all provide valuable information on the S&P 500 returns dynamics. However, to the best of our knowledge, there has been very little effort dedicated to comparing the information these datasets contain on the distribution of the S&P 500 returns and on the trajectory of their variance process. In this paper we aim to fill this gap and study the added information content of the VIX option market compared to the S&P 500 market.

Our main contribution to the empirical option pricing literature is to show that VIX options contain valuable information on the dynamic properties of S&P 500 returns, which is not spanned by S&P 500 options. We draw this conclusion from a parametric approach, using a parsimonious and flexible affine model for returns. Our results are backed by various in- and out-of-sample tests as well as an in-depth analysis of the implied variance risk premium (VRP). We argue that VIX options allow for an enhanced representation of the VRP and of its term structure, and show that the resulting VRP can be used to form trading signals and improve predictions of S&P 500 returns.

Jointly analyzing the dynamic properties and information content of the VIX and S&P 500 option markets is a challenge. Not only do we need a model that is flexible enough to simultaneously accommodate the stylized facts of both markets over time, but the empirical analysis of such highly nonlinear data poses a significant computational hurdle. We develop a time-consistent estimation procedure that permits us to extract information from a large and unbalanced panel of data and estimate the trajectories of the unobserved volatility of the S&P 500 returns. This methodology goes well beyond a simple calibration exercise as it makes it possible to reconcile time series data on the S&P 500 and VIX derivatives markets and consistently match the joint evolution of prices over time.

We model the S&P 500 returns using an affine jump-diffusion specification that belongs to the class of Duffie, Pan, and Singleton (2000). This specification features two factors driving the variance

process, and an additional factor driving the jump intensity. Its affine structure allows us to price S&P 500 and VIX derivatives in semi-closed form, which is essential to analyze the returns and volatility dynamics using a large dataset of options. It also enables us to derive the VRP in closed-form, and to conduct a thorough analysis of its dynamic behavior and term structure.

We find that VIX options contain information on the dynamics of the S&P 500 returns and their variance which is not spanned by S&P 500 options, disregarding the state of the economy. More specifically, in time of market calm, VIX options do not bring any value in estimating the current state of latent factors. However, they allow better identification of the parameters of the model, thereby providing information on the conditional distributions of the underlying returns and their variance. This translates into a better pricing of VIX options, which are not well priced when not included in the estimation dataset. This observation holds both in- and out-of-sample, and is therefore not the result of over-fitting. It also holds in times of market turmoil, but then VIX options bring value in identifying the current states of latent processes as well. Furthermore, we show that adding VIX options to the estimation allow a better representation of the term structure of variance. We synthesize the VIX index from S&P 500 options for maturities from 2 to 6 months. Our model, when estimated to a dataset which does not include VIX options, yields RMSEs which are more than twice as large as the ones obtained when VIX options are included in the estimation dataset. Our results have considerable impact in terms of pricing and risk management, which rely heavily on an accurate estimation of the conditional distributions of the underlying risk factors over different time horizons.

A by-product of our estimation is the variance risk premium (VRP), which represents the compensation investors expect to receive for bearing the risk coming from stochastic fluctuations in the variance of the returns over a given horizon. In other words, it is the expected payoff of a variance swap. By definition, it depends on the conditional expectation of the variance of S&P 500 returns. Due to the affine structure of our model, the VRP is available in closed-form, which enables us to address three highly debated questions in the recent literature: What are the main components driving the VRP? Can the VRP be used to form trading signals? Does the VRP have predictive power on S&P 500 returns?

We find that the VRP is very sensitive to jumps in the returns and their variance, in particular when the investment horizon is short. Intuitively, this is justified by the fact that a large movement in the variance process has an immediate negative impact on the payoff of a short-term variance swap. The two variance factors are shown to have different effects on the VRP. The first factor reacts swiftly to changing market conditions and captures most of the sudden variance fluctuations, especially during market turmoil. As such, its impact on the VRP dominates for short-term investments and during turmoil periods. The second factor is more persistent and captures mid- to long-term trends of the return variance. Its impact on the VRP is most important during calm market periods and for mid- to long-term investments.

Our results on the term structure of VRP complements recent findings of Gruber, Tebaldi, and Trojani (2015). In line with the latter authors, we find that that the VRP has a downward sloping term structure in times of market calm but that this is no longer true during high volatility times. During these periods, the convexity of the VRP term structure increases, i.e., the VRP becomes more negative when increasing the investment's maturity up to 3 to 4 months, and then increases. While the usual trading strategy to reap the VRP is to buy long-term variance contracts (for example variance swaps, see Filipović, Gourié, and Mancini (2016)) and sell short-term contracts, we show that the change in the term structure of VRP can be interpreted as a trading signal. Our proposed strategy achieves a Sharpe ratio of 1.02 over the turbulent period from 2006 to 2010, against 0.01 for the usual strategy.

Finally, we examine the predictive power of our model-implied VRP on S&P 500 returns. We find that our estimate over-performs by a factor close to 2 a model-free VRP which uses the ex-post realized variance as proxy for its expectation under the historical measure. However, our main finding is that the term structure of the VRP has predictive power on S&P 500 returns as well. Indeed, adding a measure of convexity of the VRP term structure (or, equivalently, of skewness), allows reaching an R^2 of 0.23 when the estimation dataset does not contain VIX options (it is composed of S&P 500 options and of the underlying S&P 500 and VIX indexes). This R^2 is around three times the one without the convexity measure. Furthermore, the R^2 increases above 0.40 when including VIX options in the estimation dataset.

As we adopt a parametric approach, our results are backed up by an extensive model specification analysis. We examine different nested models to investigate the role of the various features in explaining option prices, the risk-neutral distributions of returns, and those of the variance process. Of course, any parametric approach is bound to suffer, to a certain extent, from model misspecification. Based on likelihood criteria as well as analyses of the in- and out-of sample pricing errors, we show that the full specification of our model is needed to represent the underlying indices as well as the options on both markets.

We address the computational challenge of jointly estimating a model to two liquid option markets by designing an option pricing algorithm and a particle filter which are tailored to our problem and model specification. Estimating the dynamics of the S&P 500 returns from 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 features of the model, in particular the state-dependent jumps. To achieve this goal, we extend the Fourier 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 estimate the trajectories of unobservable processes and jumps. Accordingly, we provide an extensive toolkit for inference and diagnostics of affine option pricing models given index and option data from both the S&P 500 and VIX markets. Particle filtering techniques and more generally Sequential Monte Carlo methods have recently increased in popularity and have been used to estimate models, but most endeavors using this tool restrict their options dataset to near at-the-money options and as far as we know, none have used S&P 500 and VIX derivatives jointly.

Our work is related to several recent papers which aim 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 at pricing S&P 500 options, it fails to price VIX options. In fact, 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. Among the recent papers that have attempted to simultaneously reproduce the volatility smiles of S&P 500 and VIX options are Chung, Tsai, Wang, and Weng (2011), Cont and Kokholm (2013), Papanicolaou and Sircar (2013), and Bayer, Gatheral, and Karlsmark (2013). We build on this

literature by considering extensions of the Heston model that remain within the affine framework, but add more flexibility to the specifications used in the above mentioned papers. We use a special case of the general affine framework developed by Duffie, Pan, and Singleton (2000) that includes as sub-cases the usual extensions of the Heston model encountered in the literature, for example Bates (2000b), Eraker (2004) and Sepp (2008a).¹ In related work, Song and Xiu (2015) use a model that is similar to ours but with a different focus, and estimate marginal densities and pricing kernels of the market returns and VIX.

We also build on a literature which studies the dynamic properties of variance risk premia. Amengual (2008) uses S&P 500 options and variance swaps to infer the term structure of variance risk premia prior to the financial crisis. He finds a downward-sloping term structure of variance risk premia, which is confirmed by later studies by Andries, Eisenbach, Schmalz, and Wang (2015), solely based on S&P 500 options. Gruber, Tebaldi, and Trojani (2015) are the first to differentiate between periods of low and high volatility, and find that the slope of the VRP term structure switches sign in periods of distress. Our results extend theirs as we show that the term structure of VRP is no longer monotonic in times of high volatility: it has negative slope up to three to four months, and then positive slope. On a related note, Dew-Becker, Giglio, Le, and Rodriguez (2015) show that on average, investors do not price news about future volatility and are only willing to hedge against fluctuations in short-term realized volatility. We argue that the attitude of investors towards long-term volatility changes over time, and depends on market conditions.

Finally, our work enriches the literature on time-consistent estimation methods. These methods have been previously used to calibrate models to index returns and options. See, e.g., Bates (2000a), Pan (2002), Eraker (2004), Broadie, Chernov, and Johannes (2007), Christoffersen, Jacobs, and Mimouni (2010), Johannes, Polson, and Stroud (2009) and Duan and Yeh (2011). However, as underlined in Ferriani and Pastorello (2012), most papers filtering information from option prices rely on one option per day or a limited set of options. Limiting the amount of data results in a computationally less intensive empirical exercise, but it ignores a large part of the information present in the markets. In

¹Some studies are going in the direction of non-affine models (e.g., Jones (2003), Ait-Sahalia and Kimmel (2007), Christoffersen, Jacobs, and Mimouni (2010), Ferriani and Pastorello (2012), Durham (2013), 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.

contrast, in our particle filter estimation we fully exploit the richness of our dataset. Furthermore, we note that most if not all papers that consider S&P 500 and VIX options in their calibration exercise have restricted their analysis to a static one-day estimation. The resulting parameters might exhibit large variations when calibrating the model to different dates and therefore cannot be used to infer time series properties of returns and risk premia.²

This paper is organized as follows. In Section 2, we introduce the three-factor affine jump–diffusion framework used later in the estimation. We describe the risk premium specification and derive the expression for the VIX squared as well as the pricing formula for VIX and S&P 500 options. In Section 3, we describe our dataset. In Section 4, we detail our time series consistent estimation method. In Section 5, we discuss our estimation results and model specification analysis. Section 6 provides a thorough analysis of the VRP, the properties of its term structure and predictive power on future S&P 500 returns. Section 7 concludes.

2 Theoretical framework

We first present our modeling framework. Our model is novel and able to represent important stylized facts of S&P 500 returns which have been recently highlighted in the literature. In particular, it includes a state-of-the-art representation of the jumps, inspired from Andersen, Fusari, and Todorov (2015) and Amengual and Xiu (2015), which allows better capturing the stochastic skewness of returns and of their variance. Despite its flexibility, it is parsimonious and belongs to the affine class of models, which ensures tractability.

2.1 Model specification

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be a filtered probability space satisfying the usual assumptions, where \mathbb{P} denotes the historical measure. We consider a risk-neutral measure \mathbb{Q} equivalent to \mathbb{P} . Let $(F_t)_{t \geq 0}$ be the forward price of the S&P 500 index and $Y = (Y_t)_{t \geq 0} = (\log(F_t))_{t \geq 0}$ the returns. The dynamics

²See, e.g., Lindström, Ströjby, Brodén, Wiktorsson, and Holst (2008).

of Y under \mathbb{Q} are specified by

$$dY_t = \mu(v_{t-}, m_t, u_t)dt + \sqrt{v_{t-}}dW_t^Y + dJ_t^{Y(+)} + dJ_t^{Y(-)}, \quad (1)$$

$$dv_t = \kappa_v(m_{t-} - v_{t-})dt + \sigma_v\sqrt{v_{t-}}dW_t^v + dJ_t^{v(+)} + dJ_t^{v(-)}, \quad (2)$$

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

where W^Y , W^v , W^m and W^u are standard Brownian motions. The processes W^Y and W^v are correlated with coefficient $\rho_{Y,v}$. All other Brownian motions are independent of one and another.

The process $v = (v_t)_{t \geq 0}$ is the diffusive component of the variance of S&P 500 returns. The second variance factor $m = (m_t)_{t \geq 0}$ represents a stochastic level around which v reverts.³ We refer to it as central tendency. The two processes v and m are instantaneously uncorrelated and only interact via the drift term of v .

The processes $J_t^{Y(+)}$, $J_t^{Y(-)}$, $J_t^{v(+)}$ and $J_t^{v(-)}$ are finite activity jump processes driven by the point processes $N_t^{Y(+)}$ and $N_t^{Y(-)}$. The process $N_t^{Y(-)}$ (resp. $N_t^{Y(+)}$) counts negative (resp. positive) jumps in the returns. The jump processes are defined by

$$J_t^{Y(+)} = \sum_{i=1}^{N_t^{Y(+)}} Z_i^{Y(+)}, \quad J_t^{Y(-)} = \sum_{i=1}^{N_t^{Y(-)}} Z_i^{Y(-)}, \quad J_t^{v(+)} = \sum_{i=1}^{N_t^{Y(-)}} Z_i^{v(+)}, \quad J_t^{v(-)} = \sum_{i=1}^{N_t^{Y(+)}} Z_i^{v(-)}. \quad (4)$$

where $Z_i^{Y(+)}$, $Z_i^{Y(-)}$, $Z_i^{v(+)}$ and $Z_i^{v(-)}$ are the random jump sizes. As suggested by the price paths of the S&P 500 and VIX index, large negative (res. positive) movements in equity returns and large positive (resp. negative) movements in the variance are likely to occur at the same time. We therefore choose, in line with the literature, the same point processes $N_t^{Y(+)}$ and $N_t^{Y(-)}$ to generate jumps in the asset returns and variance process v . The leverage effect is driven by the correlation between W^Y and W^v as well as the possibility of simultaneous jumps in the returns and variance. While it is standard to model positive jumps in the returns, accounting for negative jumps has less been

³It has already been shown that at least two factors are needed to provide an accurate description of the volatility dynamics (see, e.g., Bates (2000b), Andersen, Benzoni, and Lund (2002), Alizadeh, Brandt, and Diebold (2002), Chernov, Gallant, Ghysels, and Tauchen (2003), Christoffersen, Heston, and Jacobs (2009), Egloff, Leippold, and Wu (2010), Todorov (2010), Kaeck and Alexander (2012), Bates (2012), Johnson (2012), Mencía and Sentana (2013) and Huang and Shaliastovich (2015)).

done. Amengual and Xiu (2015) show that negative jumps in volatility do occur, and are usually triggered by macroeconomic announcements. We assume that the jump intensities depend linearly on the factor levels⁴. The intensity of positive jumps in returns is denoted by $\lambda^{(+)}$, and the intensity of negative jumps by $\lambda^{(-)}$:

$$\lambda^{(+)}(v_{t-}, m_t) = \lambda_0^{(+)} + \boldsymbol{\lambda}^{(+)\top} \mathbf{X}_{t-} ; \quad \lambda^{(-)}(v_{t-}, m_t, u_t) = \lambda_0^{(-)} + \boldsymbol{\lambda}^{(-)\top} \mathbf{X}_{t-}, \quad (5)$$

where \mathbf{X} denotes the (column) vector of latent processes $\mathbf{X}_{t-} = (v_{t-}, m_t, u_t)^\top$, $\boldsymbol{\lambda}^{(+)} = (\lambda_1^{(+)}, \lambda_2^{(+)}, 0)^\top$ and $\boldsymbol{\lambda}^{(-)} = (\lambda_1^{(-)}, \lambda_2^{(-)}, \lambda_3^{(-)})^\top$. The process u drives the intensity of negative jumps in the returns (joint with positive jumps in the variance), as in Andersen, Fusari, and Todorov (2015). It has the following dynamics:

$$du_t = \kappa_u(\theta_u - u_t)dt + \sigma_u \sqrt{u_t} dW_t^u, \quad (6)$$

with W^u independent of the other Brownian motions. Intuitively, u allows to better represent the stochastic skewness of the return process. Andersen, Fusari, and Todorov (2015) find that the effect of the process u on the intensity of positive jumps in returns and on the diffusive variance is insignificant, therefore we do not incorporate it in our model.

We assume that the random jump sizes are independent and identically distributed. The jump sizes in the returns and volatility are assumed to be exponentially distributed with respective means $\mu_Y^{(+)}$, $\mu_Y^{(-)}$, $\nu_v^{(+)}$ and $\nu_v^{(-)}$. Let us define $\mathbf{Z}_i = (Z_i^{Y^{(+)}} , Z_i^{Y^{(-)}} , Z_i^{v^{(+)}} , Z_i^{v^{(-)}})^\top$, $i \in \mathbb{N}^*$. The jump sizes are characterized by their joint Laplace transform

$$\theta_Z(\phi) = \theta_{Z_1}(\phi_Y^{(+)}, \phi_Y^{(-)}, \phi_v^{(+)}, \phi_v^{(-)}) = \mathbb{E}^{\mathbb{Q}}[\exp(\boldsymbol{\phi}^\top \mathbf{Z}_1)], \quad \boldsymbol{\phi} \in \mathbb{C}^3. \quad (7)$$

⁴The specification of jumps is of importance. Todorov (2010), Todorov and Tauchen (2011) and Jacod and Todorov (2010) find striking evidence for co-jumps in S&P 500 returns and in the VIX. See also Eraker (2004), Broadie, Chernov, and Johannes (2007), Cont and Kokholm (2013). 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) find that the state dependent intensity of jumps is a desirable model feature.

The drift of the returns process can be written accordingly as:

$$\mu(v_{t-}, m_t, u_t) = -\lambda^{(-)}(v_{t-}, m_t, u_t)(\theta_Z(0, 1, 0, 0) - 1) - \lambda^{(+)}(v_{t-}, m_t)(\theta_Z(1, 0, 0, 0) - 1) - \frac{1}{2}v_{t-} \quad (8)$$

In our model, the diffusive variance of returns can in theory reach zero with positive probability as well as become negative because of the negative jumps in v . While this is certainly a drawback, the empirical literature has highlighted the need for negative jumps in volatility and shown that even with a standard Heston model with positive jumps only, imposing the Feller condition puts restrictions on the model which do not allow a satisfactory representation of the data, see, e.g., Song and Xiu (2015). To tackle this issue, we verify in the empirical part of the paper that the estimated trajectory of the process v_t never crosses the zero boundary.

This model specification implicitly defines the dynamics for the VIX. To derive its expression within our framework, we use the definition of the VIX as a finite sum of call and put prices that converges to the integral $\text{VIX}_t^2 = \frac{2}{\tau} \mathbb{E}_t^{\mathbb{Q}} \left[\int_t^{t+\tau} \frac{dF_u}{F_u} - d(\ln F_u) \right]$, where τ is 30 days in annual terms.

Proposition 2.1. *Under the model specification given in Equations (1)-(7), the VIX squared at time t can be written as an affine deterministic function of v_t and m_t :*

$$\text{VIX}_t^2 = \frac{1}{\tau} \mathbb{E}_t^{\mathbb{Q}} \left[\int_t^{t+\tau} v_u du + 2 \sum_{i=N_t^{Yv}}^{N_{t+\tau}^{Yv}} \left(e^{Z_i^Y} - 1 - Z_i^Y \right) \right], \quad (9)$$

$$= \alpha_{\text{VIX}^2} v_t + \beta_{\text{VIX}^2} m_t + \gamma_{\text{VIX}^2} u_t + \delta_{\text{VIX}^2}, \quad (10)$$

where the coefficients α_{VIX^2} , β_{VIX^2} , γ_{VIX^2} and δ_{VIX^2} are known in closed form and provided in Appendix A.

2.2 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} . We separate the total equity risk premium γ_t into a diffusive contribution, which is proportional to the variance level and represents the compensation

for the diffusive price risk, and a jump contribution reflecting the compensation for jump risk:

$$\gamma_t = \eta_Y v_{t-} + \lambda^{(+)}(v_{t-}, m_t) \left(\theta_Z^{\mathbb{P}}(1, 0, 0, 0) - \theta_Z(1, 0, 0, 0) \right) + \lambda^{(-)}(v_{t-}, m_t, u_t) \left(\theta_Z^{\mathbb{P}}(0, 1, 0, 0) - \theta_Z(0, 1, 0, 0) \right), \quad (11)$$

where $\theta_Z^{\mathbb{P}}$ denotes the joint Laplace transform of jump sizes under the historical measure \mathbb{P} . We follow Pan (2002) and Eraker (2004) and assume that the intensity of jumps is the same under \mathbb{Q} and \mathbb{P} .⁵ However, we allow the mean of the jump sizes in returns to be different under \mathbb{Q} and \mathbb{P} .

Similarly, the volatility risk premium on the two volatility factors v and m decomposes 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 $\eta_v = \kappa_v - \kappa_v^{\mathbb{P}}$. The same applies to the central tendency m , for which the coefficient is defined as $\eta_m = \kappa_m - \kappa_m^{\mathbb{P}}$. For the jump part of the volatility risk premium, we allow the mean jump sizes $\nu_v^{(+)}$ and $\nu_v^{(-)}$ to be different under \mathbb{P} and \mathbb{Q} .

In line with Andersen, Fusari, and Todorov (2015), we find in the empirical part that the trajectory of the factor u is difficult to estimate. Therefore, to not add unnecessary complexity to the model, we assume that it does not carry any risk premium.

2.3 Derivatives pricing

Within the class of affine models, option pricing is most efficiently performed using Fourier inversion techniques. As a starting point, we need the characteristic function of the underlying processes. Due to the affine property of the VIX square in Proposition 2.1, we have the following result:

Proposition 2.2. *In the two-factor stochastic volatility model with jumps defined by equations (1)-(7), the Laplace transforms of the VIX square and the S&P 500 returns are exponential affine in the*

⁵Pan (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. Our assumption artificially incorporates the jump-timing risk premium into the mean jump size risk premium.

current values of the factor processes v , m and u :

$$\begin{aligned}\Psi_{VIX_T^2}(t, \tilde{x}; \omega) &:= \mathbb{E}_t^{\mathbb{Q}} \left[e^{\omega VIX_T^2} \mid \mathbf{X}_t = \tilde{x} \right] = e^{\alpha(T-t) + \mathbf{B}(T-t)\tilde{x}}, \\ \Psi_{Y_T}(t, y, \tilde{x}; \omega) &:= \mathbb{E}_t^{\mathbb{Q}} \left[e^{\omega Y_T} \mid y_t = y, \mathbf{X}_t = \tilde{x} \right] = e^{\alpha_Y(T-t) + \mathbf{B}_Y(T-t)\tilde{x}},\end{aligned}$$

where $\mathbf{B} = (\beta, \gamma, \delta)^\top$ and $\mathbf{B}_Y = (\gamma_Y, \delta_Y, \xi_Y)^\top$ are functions defined on $[0, T]$ by the ODEs given in Appendix B. The parameter ω belongs to a subset of \mathbb{C} where the above expectations are finite.

Pricing options on the VIX poses technical difficulties that are not encountered when pricing equity options. Given a call option with strike K and maturity T on the VIX at time $t = 0$, we need to calculate

$$C(VIX_0, K, T) = e^{-rT} \int_0^\infty (\sqrt{v} - K)^+ f_{VIX_T^2}(v) dv, \quad (12)$$

where $f_{VIX_T^2}$ is the \mathbb{Q} -density of the VIX square at time $t = T$. The square root appearing in the integral as part of the payoff in (12) prevents us from using the Fast Fourier Transform of Carr and Madan (1999). We would need the log of the VIX to be affine, which is incompatible with affine models for log-returns.

However, this problem can be circumvented. Fang and Oosterlee (2008) introduce the Fourier cosine expansion to price index options on the S&P 500. We extend their method to tackle the pricing of VIX options. Our approach to pricing VIX options is comparable to the inversion performed by Sepp (2008a) and Song and Xiu (2015), but it is more parsimonious in the number of computational parameters.

Proposition 2.3. *Consider a European-style contingent claim on the VIX index with maturity T and payoff $u_{VIX}(VIX^2) = (\sqrt{VIX^2} - K)^+$. Given an interval $[a_{VIX}, b_{VIX}]$ for the support of the $VIX_T^2|_{v_0, m_0}$ density, the price $P_{VIX}(t_0, VIX_0)$ at time $t = t_0 \geq 0$ of the contingent claim is*

$$P_{VIX}(t_0, VIX_0) = e^{-r(T-t_0)} \sum_{n=0}^{N-1} A_n^{VIX^2} U_n^{VIX^2}, \quad (13)$$

where the prime superscript in the sum \sum' means that the first term $A_0^{\text{VIX}^2} U_0^{\text{VIX}^2}$ is divided by 2. The terms in the sum are defined by

$$A_n^{\text{VIX}^2} = \frac{2}{b_{\text{VIX}} - a_{\text{VIX}}} \text{Re} \left\{ \Psi_{\text{VIX}_T^2} \left(t_0, v_0, m_0; \frac{in\pi}{b_{\text{VIX}} - a_{\text{VIX}}} \right) \exp \left(-ia_{\text{VIX}} \frac{n\pi}{b_{\text{VIX}} - a_{\text{VIX}}} \right) \right\}, \quad (14)$$

$$U_n^{\text{VIX}^2} = \int_{a_{\text{VIX}}}^{b_{\text{VIX}}} u_{\text{VIX}}(v) \cos \left(n\pi \frac{v - a_{\text{VIX}}}{b_{\text{VIX}} - a_{\text{VIX}}} \right) dv. \quad (15)$$

The coefficient $A_n^{\text{VIX}^2}$ is computed using Proposition 2.2 and $U_n^{\text{VIX}^2}$ is known in closed form and given in Appendix C.

3 Data and preliminary analysis

In this section, we describe our data and point out some important characteristics of VIX options. We also perform a preliminary model selection exercise, in which we try to jointly calibrate some candidate models to the S&P 500 and VIX option markets.

3.1 Data description

Options on the VIX were introduced in 2006. Our sample period is from March 1, 2006 to October 29, 2010. The option data consist of the daily closing prices of European options on the S&P 500 and VIX, obtained from OptionMetrics. This time series includes both periods of calm and periods of crisis with extreme events. Therefore, it provides an ideal test bed for our candidate models.

Both the S&P 500 and VIX options datasets are treated following the literature, see e.g. Aït-Sahalia and Lo (1998). We only consider options with maturities between one week and one year and delete options quotes that are not traded on a given date. Then, 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. We only work with liquid out-of-the-money (OTM) options for the S&P 500

market and only with liquid call options for the VIX market. If the VIX in-the-money (ITM) call is not liquid, we use the put–call parity to infer a liquid VIX ITM call from a more liquid VIX OTM put. Finally, implied volatilities are computed considering futures prices as underlying.⁶

These adjustments leave a total of 383,286 OTM S&P 500 and 43,775 call options on the VIX, with a daily average of 327 S&P 500 options and 37 VIX options. The number of S&P 500 (VIX) options in our dataset on a given date increases with time, with around 170 (5) options at the beginning of the dataset and around 450 (70) options at the end. At the beginning of the sample, there are one or two short maturities (less than six months) available for VIX options and around six maturities for S&P 500 options, with approximately 40 S&P 500 options per maturity slice. At the end of the sample, the VIX options have around five short maturities 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 comes from the fact that VIX options only started trading in 2006. At the end of our sample, the total VIX options volume per day is about one-half the total volume of S&P 500 options traded.

3.2 Descriptive statistics

Table 1 presents the first four sample moments of the S&P 500 futures returns and VIX index levels, over two different periods of time. The first period starts in March 2006 and ends in February 2009, i.e., it spans the pre-crisis period as well as the beginning of the crisis. The second period begins in March 2009 and lasts until October 2010. For our estimation, these two periods serve as the in-sample and out-of-sample periods.

The S&P 500 log-returns exhibit a high kurtosis, especially during the in-sample period, suggesting the presence of rare and large movements. In the in-sample period, their skewness is slightly positive, but becomes negative in the out-of-sample period, due to the substantial losses made during the

⁶We remark that VIX option prices do not satisfy no-arbitrage relations with respect to the VIX index, but rather with respect to the VIX futures value. A VIX call option at time t with maturity T is an option on the volatility for the time interval $[T, T + 30d]$, where $30d$ stands for 30 days. The value VIX_t at time t is related to the volatility on the time interval $[t, t + 30d]$, which might not overlap at all with $[T, T + 30d]$.

financial crisis. The VIX index exhibits a large positive skewness and kurtosis in the in-sample period. However, in the out-of-sample period, both skewness and kurtosis decrease significantly whereas the mean increases by 45%, indicating that the values of the index are of larger overall magnitude but with less extreme values.

[Table 1 about here.]

Panel A of Figure 1 displays the joint evolution of the S&P 500 and the VIX index from 2006 to 2010. The S&P 500 returns and the VIX daily increments are highly negatively correlated (with a correlation coefficient of -0.832 over this period), which explains the popularity of VIX contracts for hedging part of the equity risk of a portfolio. Panel B represents the expected one-month forward log-returns of the S&P 500 from March 1st, 2006 to October 29th, 2010 as implied by prices of S&P 500 options and calculated following the method of Bakshi, Kapadia, and Madan (2003). The expected forward returns illustrate the variety of market situations covered by our time series. 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. From the end of 2007, they exhibit more variation and eventually turn negative. Following the bail-out of Lehman Brothers in September 2008, the expected forward returns drop below -1.5% . Then, they gradually come back and become stable mid-2009 around a slightly negative level close to -0.2% . In 2010, the sudden increase in the VIX index coincides with a further sudden drop of the expected forward returns falling to almost -0.5% .

[Figure 1 about here.]

Even though the S&P 500 and VIX markets are closely related, we emphasize that options on the VIX and S&P 500 differ substantially in their characteristics and in the information that they contain on the underlying S&P 500 returns and variances. First, S&P 500 and VIX derivatives with the same maturity contain information on the S&P 500 over different time periods. While 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 , 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 two types of options also differ in their contents. While S&P 500 options, assuming a

continuous range of traded strikes, characterize the conditional density of future S&P 500 returns, it only provides a point estimate, at each traded maturity, of the conditional variance of returns. In turn, VIX options characterize the whole conditional density of future VIX levels, and as such are likely to contain much more information on the future density of S&P 500 return variance.

Panels C and D of Figure 1 display the S&P 500 and VIX smiles on May 10, 2010. The implied volatilities (IVs) are computed using the standard Black–Scholes formula. The VIX IVs are in general substantially higher than the S&P 500 IVs. They range in our sample from 34% to 216% with an average of 80% whereas the S&P 500 IVs range from 6% to 162% with an average of 26%. The implied volatilities of S&P 500 options are in general decreasing with moneyness, which highlights the expensiveness of out-of-the money put options on the S&P 500. As these options provide investors with protection against large downward movements in returns, the negative skewness of the volatility smile reflects their risk aversion towards such movements. Due to the leverage effect, negative changes in returns are strongly related to increases in volatility, which out-of-the-money VIX call options can hedge. This explains why VIX implied volatilities tend to be positively skewed. A related quantity is 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 double that of the puts. In fact, we can observe in Panels C and D of Figure 1 that the log-moneynesses traded for S&P 500 options are mostly negative (which corresponds to OTM put options) and often positive for VIX options (OTM calls).

4 Estimation methodology

Before we bring our model to the time series of data, we carry out a joint calibration exercise using the cross-section of S&P 500 and VIX options on a particular date. This exercise gives us some guidance for model design and allows us to reduce the set of models to be estimated on a time series of options’ data. If a model is not flexible enough to jointly reproduce the implied volatility patterns of both markets on a single date, the \mathbb{Q} -dynamics of the model is not sufficiently rich to accurately price both the S&P 500 and VIX derivatives jointly, and we can safely discard this model

from further consideration. Our methodology and detailed results are reported in Appendix D. We find that irrespective of the day, the specification of our model with jumps but m and u constant (SVJ) performs as well as the specification with m stochastic and u constant (SVJ2) and the full specification with u stochastic too (SVJ3) on the S&P 500 options market, all fitting the data very well. In contrast, we find 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 and SVJ3 models satisfactorily fit both.⁷ Therefore, we conclude from our calibration exercise that we can discard the Heston model from further analysis and that jumps in the volatility are necessary.

Daily calibration is essentially a multiple curve fitting exercise which matches a model to risk-neutral distributions implied by option prices at different maturities. Some of the parameters we get from daily calibrations are unstable and vary substantially from one day to the next.⁸ To achieve a more robust estimation which is consistent with the whole time series of in-sample data, we choose a methodology based on particle filtering. A particle filter uses a time series of observable market data, called measurements, to estimate the conditional densities of unobserved latent processes such as the volatility and jump processes at every point in time during the estimation period. It can be combined with maximum likelihood estimation for parameter estimation and standard error calculations. Using a time series of S&P 500 and VIX indexes and options, we estimate both the \mathbb{P} - and \mathbb{Q} -dynamics of the model to obtain a set of model parameters that jointly prices spot and options in both markets consistently over time. The estimation is performed over the in-sample period. The out-of-sample analysis is conducted by setting the parameters equal to their in-sample estimates and running the filter on the subsequent period. Before introducing the specific filter used, we specify the discretized model and the measurement errors.

⁷Our findings are consistent with Gatheral (2008), who shows that the Heston model is incapable of reproducing the positive skew in VIX IVs, and with Sepp (2008a,b), who finds that incorporating positive jumps in the volatility dynamics into the Heston model removes this shortcoming.

⁸Parameters obtained when calibrating to daily options prices are not stable over time, as explained in Broadie, Chernov, and Johannes (2007) and Lindström, Ströjby, Brodén, Wiktorsson, and Holst (2008).

4.1 Discretized model and specification of errors

We discretize the continuous-time model on a uniform time grid composed of $M + 1$ points $t \in \{t_0 = 0, t_1 = \Delta t, \dots, t_k = k\Delta t, \dots, t_M = M\Delta t\}$, for some $M \in \mathbb{N}^*$. Since we use daily data, Δt corresponds to one day. In discrete time, the model evolves under \mathbb{P} as follows:

$$\begin{aligned} \Delta Y_t &= [-\lambda^{(-)}(v_t, m_t, u_t)(\theta_Z^{\mathbb{P}}(0, 1, 0, 0) - 1) - \lambda^{(+)}(v_t, m_t)(\theta_Z^{\mathbb{P}}(1, 0, 0, 0) - 1) - \frac{1}{2}v_t + \gamma_t]\Delta t \\ &\quad + \sqrt{v_t}\Delta W_t^{Y, \mathbb{P}} + Z_t^{Y(+), \mathbb{P}}\Delta N_t^{(+)} + Z_t^{Y(-), \mathbb{P}}\Delta N_t^{(-)}, \end{aligned} \quad (16)$$

$$\Delta v_t = \kappa_v^{\mathbb{P}} \left(\frac{\kappa_v}{\kappa_v^{\mathbb{P}}} m_t - v_t \right) \Delta t + \sigma_v \sqrt{v_t} \Delta W_t^{v, \mathbb{P}} + Z_t^{v(-), \mathbb{P}} \Delta N_t^{(+)} + Z_t^{v(+), \mathbb{P}} \Delta N_t^{(-)}, \quad (17)$$

$$\Delta m_t = \kappa_m^{\mathbb{P}} (\theta_m^{\mathbb{P}} - m_t) \Delta t + \sigma_m \sqrt{m_t} \Delta W_t^{m, \mathbb{P}} \quad (18)$$

$$\Delta u_t = \kappa_u (\theta_u - u_t) \Delta t + \sigma_u \sqrt{m_t} \Delta W_t^{u, \mathbb{P}} \quad (19)$$

where the notation ΔX_t for some process X represents the increment $X_{t_{k+1}} - X_{t_k}$ where $t = t_k \in \{t_0, \dots, t_{M-1}\}$.

As the log-returns are observable, equation (16) is the first measurement equation. The second measurement equation comes from the observation of the VIX index level with error:

$$\text{VIX}_t^2 - (\alpha_{\text{VIX}^2} v_t + \beta_{\text{VIX}^2} m_t + \gamma_{\text{VIX}^2} u_t + \delta_{\text{VIX}^2}) = \epsilon_t^{\text{VIX}}. \quad (20)$$

Jiang and Tian (2007) point to systematic biases in the calculation of the VIX index, such as model misspecification or data limitations. For example, in practice, the index is calculated using a finite number of options thereby inducing an error in the computation of the integral present in the definition. These biases are captured by the error term ϵ_t^{VIX} , which is assumed to follow a normal distribution with mean zero and variance $s > 0$.

In order to better identify the total variance of S&P 500 returns under the \mathbb{P} measure, we add a measurement equation which matches the logarithm of the daily Realized Variance (RV_t) of S&P 500 returns to the logarithm of the total spot variance under \mathbb{P} , as in Filipović, Gouriérou, and Mancini (2016). The associated measurement error ϵ_t is conditionally normally distributed with mean $\rho_\epsilon \epsilon_{t-1}$ and variance $c_0 + c_1 \text{RV}_{t-1}$. The rationale behind this component of the measurement equation is

the following. Andersen, Bollerslev, Diebold, and Ebens (2001), among others, provide empirical evidence that $\log(\text{RV}_t)$ is approximately normally distributed. The conditional mean specification of ϵ_t allows for autocorrelation in the measurement error, which can be induced by clustering of price jumps caused by persistence of the price jump intensity and/or microstructure noise in the estimates of daily realized variance. Autocorrelation in the measurement error is also reported in Wu (2011). The conditional variance specification of ϵ_t captures in a parsimonious way the heteroscedasticity of the measurement error due to the volatility of realized variance.

The daily RV is computed using tick-by-tick data from the S&P 500 futures and applying the two-scale estimator of Zhang, Mykland, and Ait-Sahalia (2005) with fast and slow time scales given by two-tick and 20-tick, respectively.

The last measurements are the prices of S&P 500 and VIX options. We assume that the option prices are observed with an error, which comes from different sources, such as bid-ask spreads and potential model misspecification, timing and processing errors. We define these errors as the relative differences between market $O_t^{\mathcal{M},Mkt}$ and model prices $O_t^{\mathcal{M},Mod}$, $\mathcal{M} \in \{\text{SPX}, \text{VIX}\}$:

$$\frac{O_{t,i}^{\text{SPX},Mod}(Y_t, v_t, m_t, \Theta, \Theta^{\mathbb{P},\mathbb{Q}}) - O_{t,i}^{\text{SPX},Mkt}}{O_{t,i}^{\text{SPX},Mkt}} = \epsilon_{t,i}^{\text{SPX},options}, \quad i = 1, \dots, N_{\text{SPX},t}, \quad (21)$$

$$\frac{O_{t,j}^{\text{VIX},Mod}(v_t, m_t, \Theta, \Theta^{\mathbb{P},\mathbb{Q}}) - O_{t,j}^{\text{VIX},Mkt}}{O_{t,j}^{\text{VIX},Mkt}} = \epsilon_{t,j}^{\text{VIX},options}, \quad j = 1, \dots, N_{\text{VIX},t}, \quad (22)$$

where $N_{\mathcal{M},t}$ is the number of contracts available in the corresponding market and the Θ 's are the sets of parameters to estimate:

$$\begin{aligned} \Theta^{\mathbb{P}} &= \{\kappa_v^{\mathbb{P}}, \kappa_m^{\mathbb{P}}, \theta_m^{\mathbb{P}}, \nu_m^{\mathbb{P}}, \nu_v^{(+)\mathbb{P}}, \nu_v^{(-)\mathbb{P}}, \mu_Y^{(+)\mathbb{P}}, \mu_Y^{(-)\mathbb{P}}, \eta_Y\} \\ \Theta &= \{\kappa_v, \kappa_m, \theta_m, \nu_v^{(+)}, \nu_v^{(-)}, \mu_Y^{(+)}, \mu_Y^{(-)}\} \\ \Theta^{\mathbb{P},\mathbb{Q}} &= \{\lambda_0^{(-)}, \lambda_1^{(-)}, \lambda_2^{(-)}, \lambda_3^{(-)}, \lambda_0^{(+)}, \lambda_1^{(+)}, \lambda_2^{(+)}, \sigma_m, \sigma_v, \sigma_u, \kappa_u, \theta_u, \rho_{Yv}\}. \end{aligned}$$

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

$$\epsilon_{t,i}^{\text{SPX},options} \sim \mathcal{N}(0, \sigma_{\epsilon_{t,i}^{\text{SPX}}}^2), \quad \epsilon_{t,j}^{\text{VIX},options} \sim \mathcal{N}(0, \sigma_{\epsilon_{t,j}^{\text{VIX}}}^2), \quad (23)$$

The variance of the errors is

$$\sigma_{\epsilon_{t,i}^{\text{SPX}}}^2 = \exp\left(\phi_0 \cdot \text{bid-ask spread}_i + \phi_1 \left| \log\left(\frac{K_i}{F_t^{\text{SPX}}(T_i)}\right) \right| + \phi_2(T_i - t) + \phi_3\right), \quad (24)$$

$$\sigma_{\epsilon_{t,j}^{\text{VIX}}}^2 = \exp\left(\psi_0 \cdot \text{bid-ask spread}_j + \psi_1 \left| \log\left(\frac{K_j}{F_t^{\text{VIX}}(T_j)}\right) \right| + \psi_2(T_j - t) + \psi_3\right), \quad (25)$$

with ϕ_i and ψ_i in \mathbb{R} , $i \in \{0, \dots, 3\}$.⁹

4.2 Particle filter

At every period $t = t_n$, the measurement vector y_t collects observed market prices. By $y^t = (y_{t_0}, \dots, y_{t_n})$, we denote all the observations available up to time t . The filtering problem consists of recursively approximating the distribution of the latent state L_t ,

$$L_t = \left\{ v_t, m_t, u_t, \Delta N_t^{(+)}, \Delta N_t^{(-)}, Z_t^{Y^{(+)}} , Z_t^{Y^{(-)}} \right\}, \quad (26)$$

conditional on y^t . Particle filters are perfectly adapted to our problem: They can handle observations that are nonlinear functions of latent variables as well as equations with non-Gaussian innovations.

There are many types of particle filters. We use the Auxiliary Particle Filter (APF) proposed by Pitt and Shephard (1999). Compared to more basic particle filters, such as the Sampling Importance Resampling (SIR) filter, the APF is better suited to detect jumps, whereas the SIR filter faces sample impoverishment leading to potential particle degeneracy. Both filters are described in Johannes, Polson, and Stroud (2009) for filtering latent factors from returns in a Heston model with jumps in returns.

We develop an extension of their algorithm that is able to handle more data (the VIX market data on top of the S&P 500 market data) as well as the second volatility factor m , the third factor for jumps u_t and the volatility jumps. The likelihood estimation and particle filter are presented in detail in Appendix E. In particular, we use the weighted likelihood method of Hu and Zidek (2002), in order

⁹The fact that the option pricing errors are normally distributed does not constitute a restriction. The reason is that the errors are heteroscedastic and the coefficients generating the heteroscedasticity are driven by the data, i.e., we optimize over the parameters $\{\phi_i, \psi_i\}_{0 \leq i \leq 3}$.

to assign comparable weight to S&P 500 and VIX options. Furthermore, we performed additional data treatments for S&P 500 and VIX options before running the particle filter. They are described in Appendix F.

5 Estimation results with and without VIX options

This section analyzes the benefits of including the different features of the full model specification, with and without VIX options in the estimation dataset.

5.1 Terminology

We denote by \mathcal{D}_1 the dataset containing S&P 500 options, S&P 500 returns and their daily RV and VIX levels, and by \mathcal{D}_2 the dataset containing VIX options in addition the components of \mathcal{D}_1 . Furthermore, the full specification of model (1)-(6) is called SVJ3 (Stochastic Volatility model with Jumps and 3 latent factors). In this model and to improve identification, we impose θ_u to 1 and $\lambda_0^{(-)} = 0$. The following sub-specifications of the SVJ3 model are examined: (i) a specification where m is a constant and the intensity of jumps does not load on u , denoted SVJ; (ii) a specification where m is stochastic but the intensity of jumps does not load on u , denoted SVJ2.

5.2 Likelihood criteria and parameter estimates

Before we discuss in detail our parameter estimates, we would like to stress the fact that options are crucial for identifying the parameters of our model. Even when estimating highly restricted sub-specifications of our full specification to a dataset without options (with the underlying S&P 500 and VIX indices only), we find that the resulting estimates of the \mathbb{Q} -parameters (and some of the \mathbb{P} -parameters) have extremely large standard errors, typically four to five times larger than the ones obtained with datasets containing options. This problem can partly be resolved by extending the in-sample time period, leading to a more accurate estimation of the \mathbb{P} -parameters, but not of the \mathbb{Q} -parameters. Therefore the VIX index does not contain sufficient information to identify the

\mathbb{Q} -dynamics of S&P 500 returns, as has been argued in, e.g., Duan and Yeh (2010, 2011).¹⁰

Table 2 reports the log-likelihood across the model sub-specifications and the values of the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC), with both estimation datasets.

[Table 2 about here.]

Inspection of these values for \mathcal{D}_2 suggests that the full SVJ3 specification is substantially superior to all nested sub-specifications examined. When VIX options are not part of the estimation dataset (\mathcal{D}_1), the SVJ3 model only slightly out-performs the SVJ2 model, where the intensity of jumps does not load on the additional factor u . Given that this factor controls for simultaneous jumps in the S&P 500 and VIX index, VIX options contain information that helps identifying the parameters driving its dynamics. This justifies why the difference in performance between the SVJ2 and SVJ3 models is larger when VIX options are included in the estimation. Other sub-specifications restricting m to a constant, or without jumps, significantly under-perform the SVJ2 and SVJ3 models for both \mathcal{D}_1 and \mathcal{D}_2 . A less flexible jump size specification where positive jumps in the returns are modeled together with negative jumps using a normal distribution also under-performs our specification significantly.

Table 3 presents the point estimates and standard errors resulting from the estimation of the SVJ3 model to datasets \mathcal{D}_1 and \mathcal{D}_2 .

[Table 3 about here.]

The estimated parameters driving the two variance processes allow identifying very different roles. Indeed, v has a high volatility parameter σ_v ranging from 0.57 to 0.76 depending on the model specification, with small standard error. Besides, it has a high speed of mean reversion under both measures, implying a half life around 33 days under \mathbb{P} and 48 days under \mathbb{Q} . In contrast, m has a volatility parameter σ_m around 10% regardless of the estimation dataset and chosen model, also with small standard error. Its speed of mean reversion is difficult to estimate precisely but ranges between 0.08 and 0.24, leading to a half life to 3 to 7 years depending on the estimation. We can interpret the process v as a factor representing erratic short-term fluctuations of the variance, whereas the process m is persistent and captures smoother medium- to long-term trends.¹¹ The estimated volatility of

¹⁰Results have not been reported for space constraints but are available upon request.

¹¹Indeed, under the assumption that jumps have a minor impact on this expectation compared to the drift term, i.e.

the jump process u , σ_u , is close to 0.3 which suggests that u is not as volatile as v but also not as persistent as m . Its quite high speed of mean reversion, corresponding to a half life between 60 and 75 days, indicates that it captures punctual events.

Not surprisingly, we find a prominent leverage coefficient ρ_{Yv} across all models and datasets.

In line with the literature, the equity risk premium coefficient η_Y is positive across all models and datasets, which implies a positive diffusive equity risk premium. The diffusive part of the volatility risk premium $\eta_v = \kappa_v - \kappa_v^{\mathbb{P}}$ is always found negative, which supports the existence of a negative variance risk premium.

Jump size estimates under \mathbb{Q} are statistically significant, with a negative part which captures the value given to large and rare events (around -11% on average). It is much more negative than the average negative jump estimate under \mathbb{P} , around -4%. This is due to investors' risk aversion to crashes and yields a non-zero negative jump risk premium. In contrast, positive jumps have comparable means under both measures. Similarly to negative jumps in returns, positive jumps in volatility are priced, with an average size of around 6% under \mathbb{Q} and 2 to 4% under \mathbb{P} . Negative jumps have a zero risk premium by assumption, but we find that their estimated size is significantly different from zero. In fact, imposing it equal to zero yields a substantial loss of likelihood.

The intensity of jumps is driven mainly by the short-term volatility process v , and to a smaller extent on the long-term process m .

$\kappa_v^{\mathbb{P}} \gg \lambda_k^{(+/-)} \nu_v^{\mathbb{P}}$ (inequalities satisfied by our parameter estimates), the conditional expectation of the variance $\mathbb{E}_t^{\mathbb{P}}[v_T]$ can be written as:

$$\mathbb{E}_t^{\mathbb{P}}[v_T] \approx \theta_m^{\mathbb{P}} \frac{\kappa_v}{\kappa_v^{\mathbb{P}}} + c \cdot e^{-\kappa_m^{\mathbb{P}}(T-t)}(m_t - \theta_m^{\mathbb{P}}) + e^{-\kappa_v^{\mathbb{P}}(T-t)}(v_t - c \cdot m_t + c \cdot \frac{\kappa_m^{\mathbb{P}}}{\kappa_v^{\mathbb{P}}} \theta_m^{\mathbb{P}}),$$

for a constant $c = \frac{\kappa_v}{\kappa_v^{\mathbb{P}} - \kappa_m^{\mathbb{P}}}$. As $\kappa_v^{\mathbb{P}} \gg \kappa_m^{\mathbb{P}}$, the coefficients in front of v_t decays much faster than the one in front of m_t . For $T - t$ equal to three months, $e^{-\kappa_v^{\mathbb{P}}(T-t)}$ is around 0.16, but goes down to 0.03 for six months, and is of order of magnitude of 10^{-3} for a year. In contrast, $e^{-\kappa_m^{\mathbb{P}}(T-t)}$ is around 0.90 for $T - t$ equal to six months, and as high as 0.80 for one year. Therefore the deviation of m_t relative to its long-term mean drives the medium- to long-term expectation of the variance.

5.3 Filtered trajectories

Figure 2, Panel A displays the filtered trajectory of the volatility process v_t resulting from estimating the SVJ3 model to \mathcal{D}_2 . Until mid-2007, the volatility remains at low levels, oscillating around 10%. When the crisis of the quant-strategy hedge funds starts in the summer of 2007, the process v gradually goes up to stabilize around 20%. The bankruptcy of Lehman Brothers in September 2008 triggers a sudden increase in v to more than 80%, closely followed by a decrease and a subsequent increase when the House of Representatives rejects the Troubled Asset Relief Program at the end of September. The volatility then reverts back to around 40%, to increase again following the distress of Bank of America in 2009. It gradually goes back to a level that is close to its initial level, to reach around 10% early 2010. The Greek financial crisis then leads to new increases in volatility, with a peak around 50% in mid-2010, and a reversion to around 10% following the agreement on a rescue plan. Note that v never touches or crosses the zero boundary. Panel B represents the difference between the estimated variance using \mathcal{D}_2 and \mathcal{D}_1 as estimation datasets. The estimated trajectory is consistent throughout the two datasets in times of market calm, and yields slightly different estimates during market turmoil. The maximum difference is around 10%. These differences can be explained as follows: first, the model is likely not able to fully represent both markets during high volatility times; second, the two markets are subject to different frictions, e.g., in terms of liquidity, or constraints and information sets of the traders involved, introducing inconsistencies between prices. Similarly, panel C represents the filtered trajectory of the stochastic central tendency m resulting from estimating the SVJ3 model to \mathcal{D}_2 . The process m is overall more persistent than the process v , in line with the parameter estimates found. It starts increasing in mid-2007 from around 1% to around 5%. It stabilizes and oscillates around that level for about a year, until September 2008 when it increases gradually again, to reach a level close to 10% at the beginning of 2009. This increase is followed by a gradual decrease until beginning of 2010, where m is around 5% again. The high volatility due to the Greek crisis in 2010 translates into a slight increase of m from 5 to 7%. While the process v reaches levels which are close to its initial level at the beginning of 2010, m stabilizes at a level which is about 7 times its initial level. Panel D represents the difference in the estimated m trajectory when removing VIX options from the estimation dataset. The trajectory does not change

dramatically, with a maximum difference of around 4% around the crisis period.

[Figure 2 about here.]

In Figure 3, Panel A represents the filtered trajectory of the jump intensity factor u , with and without VIX options. The two processes are almost symmetric about 1. This is explained by the fact that the level of mean reversion of the u factor is imposed to be 1. As the estimation with \mathcal{D}_2 yields an estimate of $\lambda_1^{(-)}$ which is much higher than the estimation with \mathcal{D}_1 , the intensity of jumps as implied by the model becomes too high and is not supported anymore by the data when v is at its peak. Similarly, with \mathcal{D}_1 , the low estimated value of $\lambda_1^{(-)}$ implies that the model-implied intensity of jumps is not high enough for the data. The factor u therefore correct for the non-linearity in v of the intensity of jumps. It is only significantly different from its long term mean 1 in times of crisis. Panel B represents the recovered jump sizes in v , when the estimated probability of jumps is larger than 10%. Note that no negative actual jump in the volatility is filtered. The data indicates that negative jumps in volatility are likely to occur with a small probability, but even during the crisis period this fact has not happened. Positive jumps in the variance have been filtered during the crisis, with and without VIX options. These jumps tend to be large, even more when VIX options are included in the estimation; the maximum jump comes close to 20%.

[Figure 3 about here.]

5.4 Pricing errors

Tables 4 and 5 present the Root Mean Square Relative Errors (RMSREs) of our model implied volatilities for S&P 500 and VIX options, in- and out-of-sample, using datasets \mathcal{D}_1 and \mathcal{D}_2 . Results are reported for the SVJ3 model. Options are sorted into buckets by moneyness and maturity.

Inspection of the pricing errors of S&P 500 options reveals that RMSREs are only marginally smaller with \mathcal{D}_1 than with \mathcal{D}_2 (0.096 versus 0.099 over all S&P 500 options in the in-sample period). This observation is valid both outside and during the financial crisis and suggests that VIX options, in-sample, do not contain information that is conflicting with the information already spanned by

S&P 500 options. In fact, Root Mean Square Errors (RMSEs)¹² even decrease from \mathcal{D}_1 to \mathcal{D}_2 , from 0.041 to 0.036.¹³ The improvement is most noticeable for deep OTM put options (RMSEs decrease from 0.091 to 0.069), OTM put options (0.022 versus 0.03) and long-term options (0.026 versus 0.034), suggesting that VIX options help identify the model parameters driving the prices of these options. This is consistent with the intuition that VIX options help pinpoint stark volatility increases (modelled with the jumps) and provide information on the long-term behavior of variance. Out-of-sample pricing errors on S&P 500 options confirm this intuition, as including VIX options allows to slightly improve both RMSREs and RMSEs overall, the strongest improvement being for OTM put options.

A comparison of pricing errors on VIX options with \mathcal{D}_1 and \mathcal{D}_2 uncovers striking results: VIX options are not well priced when they are not included in the estimation dataset. This is true in- and out-of sample, outside and during the crisis period. In-sample, the RMSRE over all VIX options decreases from 0.603 to 0.423 from \mathcal{D}_1 to \mathcal{D}_2 , and the decrease is observed throughout all maturity and moneyness buckets. The corresponding RMSE goes from 0.490 to 0.423. Out-of-sample, the numbers are comparable, with a decrease of the RMSRE from 0.653 to 0.490, and a decrease of the RMSE from 0.452 to 0.365. Note that the out-of-sample pricing errors are smaller than the in-sample errors. This is due to the fact that the main part of the financial crisis occurs during the in-sample period. Excluding options after September 2008 from the calculation of pricing errors yields an RMSRE of 0.528 with \mathcal{D}_1 versus 0.354 with \mathcal{D}_2 , and RMSEs of 0.375 and 0.255 respectively.

[Table 4 about here.]

[Table 5 about here.]

In unreported results¹⁴, we find that including a stochastic central tendency in the model significantly improves the pricing of long-term options and the representation of the tails of the distributions of

¹²RMSEs emphasize expensive options, i.e., options which are closer to the ATM level. For this reason we report RMSREs. RMSEs are available upon request.

¹³Our results are comparable to those obtained by Andersen, Fusari, and Todorov (2015) who fit a three factor volatility model to S&P 500 options. They consider IV RMSEs as the distance to minimize and obtain an RMSE of 1.7%. Excluding the financial crisis from the calculation of the RMSEs, we obtain RMSEs of 1.4%, and 3.6% when including the crisis period. In addition, their model has an additional volatility factor hence adding flexibility to fit more data.

¹⁴Pricing errors for sub-specifications of the SVJ3 model are available upon request.

the returns (OTM puts and calls on the S&P 500). Furthermore, jumps improve the representation of the right tail of the variance distribution (OTM calls on the VIX) as well as of the short-term options.

5.5 Representation of the variance term structure

In order to further assess the model's ability to reproduce the variance of S&P 500 returns, we test whether it can price claims on the variance which are not included in the estimation dataset. While the VIX measures the expectation of future variance over a 30-day horizon, it is straightforward to construct a similar index for other maturities, using S&P 500 options with the chosen maturities. In fact, in 2007 the CBOE started calculating the 3-month VXV index, and in 2013 the 9-day VXST index. We build indices for maturities 2, 3, 5 and 6 months and verify that our model can reproduce them. The RMSE on the VIX during the in-sample period is 2.6%, and goes down to 1.6% when excluding the data after September 2008 from the calculation. It is equal to 2% in the out-of-sample period. The VIX index, besides, is as well fitted when including VIX options in the estimation dataset as without them. Synthesizing a VIX index for other maturities generates a small error, which depends on the amount of traded options for the considered maturities. As there are less traded options for longer time-to-maturity options, we expect the 6-month VIX to be less precisely synthesized than the 30-day VIX index. Using \mathcal{D}_2 as estimation dataset, we obtain RMSEs on the synthesized 2, 3, 5 and 6-month VIX indexes which are equal to, respectively, 3.4%, 3.1%, 2.8% and 2.8%, over the entire time period considered (in- and out-of-sample). In contrast, using \mathcal{D}_1 as estimation dataset, we obtain RMSEs of 2.3%, 4.5%, 6.8% and 7.8%, which provides striking evidence that VIX options play a fundamental role in identifying the parameters of the model that drive the variance term structure.

6 Variance risk premium

6.1 Closed-form expression

Following Bollerslev and Todorov (2011), we define the annualized integrated variance risk premium (VRP) as

$$\text{VRP}(t, T) = \frac{1}{T-t} \left[\mathbb{E}_t^{\mathbb{P}} (QV_{[t, T]}) - \mathbb{E}_t^{\mathbb{Q}} (QV_{[t, T]}) \right],$$

where $QV_{[t, T]}$ denotes the quadratic variation of the log price process, which is the sum of the integrated variance of the returns and the squared jumps in the time interval considered:

$$QV_{[t, T]} = \int_t^T v_s ds + \sum_{i=N_t^{(+)}}^{N_T^{(+)}} (Z_i^{(+)})^2 + \sum_{i=N_t^{(-)}}^{N_T^{(-)}} (Z_i^{(-)})^2$$

The VRP represents the expected payoff when buying a variance swap at time t with maturity T . We can further decompose the IVRP into a continuous and a discontinuous part:

$$\text{VRP}(t, T) = \text{VRP}^c(t, T) + \text{VRP}^d(t, T)$$

with

$$\begin{aligned} \text{VRP}^c(t, T) &= \frac{1}{T-t} \left[\mathbb{E}_t^{\mathbb{P}} \left(\int_t^T v_s ds \right) - \mathbb{E}_t^{\mathbb{Q}} \left(\int_t^T v_s ds \right) \right], \\ \text{VRP}^d(t, T) &= \frac{1}{T-t} \left[\mathbb{E}_t^{\mathbb{P}} \left(\sum_{i=N_t^{(+)}}^{N_T^{(+)}} (Z_i^{(+)})^2 + \sum_{i=N_t^{(-)}}^{N_T^{(-)}} (Z_i^{(-)})^2 \right) - \mathbb{E}_t^{\mathbb{Q}} \left(\sum_{i=N_t^{(-)}}^{N_T^{(-)}} (Z_i^{(-)})^2 + \sum_{i=N_t^{(-)}}^{N_T^{(-)}} (Z_i^{(-)})^2 \right) \right]. \end{aligned}$$

Each part can also be decomposed into a linear contribution from m_t and another one from v_t , given that the expectations of the integrated latent factors are affine in their current values, see Appendix G.

6.2 Estimated VRP

We showed in Section 5 that including VIX options in the estimation dataset allowed achieving a better representation of the conditional distribution of future variance. As the VRP is based on the expectation of future variance, we analyze its properties using dataset \mathcal{D}_2 . We estimate the VRP for investments with different times-to-maturity varying from 1 week to 1 year.

In Figure 4, Panel A, we plot the evolution of the VRP over time, with its 90% confidence interval. In line with the literature, we find that the VRP is negative. At the beginning of our sample period, it is fairly stable, around -1%, meaning that investors are willing to pay 1% of their notional per year to be hedged against variance fluctuations, when investing in a 6-month contract. From mid-2007, the VRP reacts to the slight increase in volatility, goes down and stays around -2% until Lehman Brothers' bankruptcy. This event triggers a sharp drop to almost -12%, followed by a recovery period bringing the VRP back to -2%. The Greek crisis prompts a second drop to -5%, followed by a second recovery period. In contrast with the volatility factor, the VRP never goes back to its initial level of 1%. Panel B represents the VRPs for different times-to-maturity, from 1 month to 1 year. The sign and shape of the VRPs are consistent across maturities. Before September 2008, the VRPs range between 0 and -2.5%, all of them being relatively close to one another. In September 2008, they all drop simultaneously but then recover at different paces: The shorter-term VRP recover faster than the longer-term VRP. In Panels C and D of Figure 4, we decompose the 1-week and 6-month VRPs into their continuous and discontinuous parts. The discontinuous component of the VRP dominates for shorter maturities, indicating that including jumps helps to represent the shorter end of the variance premium's term structure. In fact, the jump part of the VRP represents about two thirds of the total VRP for a 1-week investment. At the long end, the effect of a jump in the variance process is likely to be dampened by the reversion of the variance to its long-term mean, which justifies why jumps only have a moderate impact on the 6-month VRP. Finally, Panels E and F represent the contributions of the v and m factors to the VRP. The contribution of m to the VRP is negligible for short maturities but plays a substantial role for maturities larger than three months, which explains why the 1-week VRP recovers much faster than the 6-month VRP after the volatility peak. The central tendency m plays an important role in setting the level of the continuous VRP, especially for

mid- to long-term investments, but it becomes secondary during market turmoil. The VRP is then dominated by the impact of v .

[Figure 4 about here.]

Both the variance jumps and the stochastic central tendency play a crucial role in the VRP. While the jumps help represent the short end of the variance term structure, the central tendency has a large impact on the mid- to long-term VRP, especially during market calm. In times of financial crises, the impact of the process v dominates.

6.3 Term structure of the variance risk premium

While the negativity of VRP is well established, the term structure of VRP has been subject to scrutiny very recently. Indeed, in contrast to most papers finding a VRP that is more negative when time-to-maturity increases (e.g., Amengual (2008)), Gruber, Tebaldi, and Trojani (2015) find instead that the slope of the term structure switches sign in times of market distress. Such phenomenon has already been uncovered for the equity risk premium, see van Binsbergen, Hueskes, Koijen, and Vrugt (2013), and is intuitively appealing from an economic perspective. Indeed, in low volatile periods, investors are likely to require a larger compensation for an investment which has a longer time-to-horizon, as the probability that markets enter into distress before maturity increases with time-to-maturity. In turn, in highly volatile periods, and due to the mean-reverting nature of volatility, increasing the time-to-maturity of an investment also increases the probability that volatility will go down before maturity of the investment. This would justify a lower (in absolute value) VRP for longer time-to-maturity investments in times of market distress.

Panel A of Figure 5 plots the difference between the 3-month and the 1-year VRP over time. This difference is generally positive, meaning that the 1-year VRP is more negative than the 3-month VRP. During the recession (grey part of the graph), this relation is inversed and the difference becomes highly negative. In terms of modelling, we have seen that there is a large premium for jump variance, which represents about two times the premium for continuous variance fluctuations for a one-week investment. This premium disappears when increasing time-to-maturity of the investment. Panels

B and C of Figure 5 represent the continuous and jump contributions in our estimation and show that both contributions of the VRP to the 3-month minus 1-year difference are responsible for the change in sign.

[Figure 5 about here.]

Figure 6 plots the difference between the 3-month and the 6-month VIX, constructed from S&P 500 options, following the procedure used in Section 5.5. It indicates that the slope of the term structure of the \mathbb{Q} -expectation of total variance (up to an adjustment due to jumps) becomes strongly positive during the recession, driving the slope of the term structure of the variance risk premium down. Note that this observation is completely model-free, suggesting that our results are not driven by model mis-specification. The switch in the slope of term structure of VRP has strong economic implications. Indeed, in terms of investment strategy, it implies that it is no longer profitable, on average, to sell claims on variance with a longer time-to-maturity, and hedge part of the exposure by buying claims with a shorter time-to-maturity. Such investment has been shown to be dynamically optimal in Filipović, Gourié, and Mancini (2016), using variance swaps.

[Figure 6 about here.]

Let us now analyze whether the term structure of VRP can be used as trading signal. In the spirit of Dew-Becker, Giglio, Le, and Rodriguez (2015), we define a squared VIX forward as a claim on the future variance as follows:

$$\text{VIX}_{t,T_1,T_2}^2 = \text{VIX}_{t,T_2}^2 - \text{VIX}_{t,T_1}^2.$$

In other words, VIX_{t,T_1,T_2}^2 represents the price of a portfolio that is long variance with maturity T_2 and short variance with maturity T_1 . Up to a jump adjustment, it can be convenient to think of such portfolio as being long a variance swap with maturity T_2 and short a swap with maturity T_1 . Let us assume that there exists an index tracking future variance as defined above. The annual Sharpe ratio of a short investment in the $\text{VIX}_{t,3\text{ months},6\text{ months}}^2$ from 2006 until 2010 is equal to 1%. Interpreting a change of sign in the slope of the VRP term structure as a warning, and leaving the investment on hold until the slope switches sign again, yields an annual Sharpe ratio of 0.82. Switching position from selling future variance into buying future variance whenever the slope of

the VRP term structure is negative still enhances the Sharpe ratio, which reaches 1.18. The sign of the VRP slope can therefore be interpreted as a trigger indicating that selling future variance is no longer profitable in times of market turmoil. These results are robust to other values of T_1 and T_2 . For example, the previously reported Sharpe ratios become respectively 0.01, 0.55 and 1.02 if using $T_1 = 2$ months instead of 3 months.

Another property of the VRP term structure which has been recently investigated in the literature is its convexity (or concavity, in papers which define the VRP as the \mathbb{Q} -expectation of variance minus the \mathbb{P} -expectation). See, e.g., Andries, Eisenbach, Schmalz, and Wang (2015) and Dew-Becker, Giglio, Le, and Rodriguez (2015). The latter argue that shocks in the future variance (beyond the first months) are not priced, and that only transitory shocks in the short-term variance are priced. This theory is consistent with a term structure of VRP which is steep for short-term maturities, and then flattens out for longer-term maturities. It is also in line with the very small Sharpe ratio obtained when investing in the $VIX_{t,3\text{ months},6\text{ months}}^2$. Our Panel A of Figure 7 represents the term structure of VRP on different dates and shows that the level of skewness and convexity of the VRP term structure is time-dependent. Before Lehman Brothers filed for bankruptcy, the term structure of VRP is decreasing and quite flat after the 3-month horizon, corroborating the findings of Dew-Becker, Giglio, Le, and Rodriguez (2015). However, in October 2008, the shape of the VRP term structure radically changes. Not only is it no longer flat for long horizons, but it exhibits a sharp decline up to a horizon of around 3 months, reaches its minimum and then increases almost as sharply for longer-term investments. The inversion in the term structure of VRP is only happening for mid- to long-term investments. Intuitively, this finding is consistent with the belief that the situation could get worse on the short-term, i.e. in the next two to three months, but will eventually get better on the long-term. As a result, the VRP term structure exhibits a large convexity. Panel B of Figure 7 represents an estimate of the higher-order moments of the VRP term structure. The short-term skewness is measured as the difference between the 1-month and the 3-month VRP, while the long-term skewness represents the 3-month VRP minus the 1-year VRP. The convexity is measured as the sum of the 1-month and the 6-month VRP minus two times the 3-month VRP. The graph shows that the short-term skewness of the VRP term structure varies together with its convexity, in line with what is illustrated in Panel A. The long-term skewness varies in the inverse direction. In

fact, the three time-series are almost perfectly correlated, with a correlation above 0.96. In contrast to Dew-Becker, Giglio, Le, and Rodriguez (2015), we therefore find that investors care about shocks in future variance, but that the way they care differs across time.

[Figure 7 about here.]

6.4 Return predictability

Bollerslev, Tauchen, and Zhou (2009), using the difference between implied and realized variances as a proxy for the VRP, find that it has predictive power on future returns, large premia predicting high future returns. This result has been confirmed in subsequent papers.¹⁵ We analyze it using our estimates of the VRP. It does not present the disadvantage of the usual model-free estimators that approximate the non-observable expectation of the realized variance under \mathbb{P} by its realized value over a past interval. But at the same time, it has the drawback of being model-dependent. Panel A of Figure 8 represents the adjusted R^2 when regressing returns over the interval $[t, t + \tau]$, with τ between 1 month and 1 year, on different predictors. The first predictor is the difference between the VIX and the (past) 1-month realized variance used by Bollerslev, Tauchen, and Zhou (2009). The second predictor is our estimate of the 1-month VRP. Its performance is comparable to the one of the first predictor below the 2-month horizon, with negligible predictive power. Above the 2-month horizon, the R^2 increases for both predictors and our estimate of the VRP outperforms the first predictor. We include another predictor in the regression, which contains the residuals of short-term skewness of the VRP term structure, regressed on the VRP. Interestingly, the R^2 strongly increases for predictions over all horizons. For 3-month returns, the R^2 of the third regression is about five times the one of the first regression. These results indicate that not only does the VRP have predictive power on returns, but so does its term structure and its higher-order moments. Similarly to Bollerslev, Tauchen, and Zhou (2009), the R^2 increases with time-to-maturity, to reach its maximum around 3 months, and then goes down. The beta of the VRP is negative whenever the term structure component is included, confirming the usual results: the larger the VRP (in absolute value), the larger future returns on average. In contrast, the beta of the skewness variable is always

¹⁵See, among others, Drechsler and Yaron (2011), Bekaert and Hoerova (2014), Kelly and Jiang (2014) and Bali and Zhou (2016).

positive, indicating that a larger short-term skewness goes together with higher future returns on average. Panel B of Figure 8 compares the R^2 of regressions using the estimated VRP from \mathcal{D}_1 , which does not contain VIX options, and \mathcal{D}_2 which does. When using only the VRP as regressor, the R^2 is similar across maturities in both cases. However, when including the skewness of the VRP term structure in the regression, the R^2 using \mathcal{D}_2 increases substantially, more than the one using \mathcal{D}_1 : it is about twice as large over all horizons. These results confirm that VIX options contain information on the term structure of the VRP which is not spanned by S&P 500 options, and has predictive power on future returns over all horizons considered.

[Figure 8 about here.]

7 Conclusion

In this paper, we carry out an extensive empirical investigation of the information contained in VIX options on the dynamics of S&P 500 returns and their variance process. We estimate various specifications of a flexible affine model using two datasets: the first dataset contains a time series of S&P 500 and VIX indexes as well as S&P option prices, and the second dataset contains VIX options in addition to the first dataset. We do not restrict the moneyness and maturity of the options considered, to benefit from their information contents on the distribution of S&P 500 returns and their variance. Instead of a step-wise estimation, we depart from most of the literature and estimated the historical and the risk-neutral parameters jointly, in a single step. Our maximum likelihood estimation procedure is based on particle filtering.

We find that VIX options contain information on the dynamics properties of S&P 500 returns, which is not spanned by S&P 500 options. This conclusion is based on several observations. First, when VIX options are not included in the estimation dataset, they are not well priced. Including them in the estimation dataset not only corrects this in both in- and out-of-sample period, but even yields improvements of the fitting of some S&P 500 options: in-sample, deep OTM put options and long-term options are better priced and so are all put options out-of-sample. Second, including VIX options in the estimation dataset allows considerable improvements in representing the variance term

structure. We replicate the construction of the VIX index for maturities 2, 3, 5 and 6 months. We find that for maturities larger than 3 months, the model yields RMSEs which are twice as large when VIX options are not in the estimation dataset as when they are part of it.

We perform a thorough analysis of the variance risk premium (VRP) and of its term structure. In line with the literature, we find that the VRP, in absolute value, is more negative for longer time-to-maturity investments in low volatility periods. However, in high volatility periods, we uncover a non-monotonic term structure of VRP, which reaches its optimum around a 3-month maturity. This finding complements the recent results of Gruber, Tebaldi, and Trojani (2015) and Dew-Becker, Giglio, Le, and Rodriguez (2015). We illustrate the economic implications of our results by testing a simple trading strategy which sells long-maturity variance and buys short-maturity variance, switching the sign of positions when an inversion of the term structure occurs. Our strategy allows reaping the variance risk premium disregarding market conditions, and reaches a Sharpe ratio of 1.18, against 0.01 for a strategy which does not react to changes in the VRP term structure. Finally, we study the predictive power of our model-implied VRP and its term structure on future S&P 500 returns, and find that adding a term structure component substantially improves the predictive power of the VRP. This holds especially true when VIX options are included in the estimation dataset.

References

- Aït-Sahalia, Y., M. Karaman, and L. Mancini, 2012, “The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis,” working paper, Princeton University and Swiss Finance Institute.
- Aït-Sahalia, Y., and R. Kimmel, 2007, “Maximum Likelihood Estimation of Stochastic Volatility Models,” *Journal of Financial Economics*, 83(2), 413–452.
- Aït-Sahalia, Y., and A. Lo, 1998, “Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices,” *Journal of Finance*, 53(2), 499–547.
- Alizadeh, S., M. W. Brandt, and F. X. Diebold, 2002, “Range-based estimation of stochastic volatility models,” *Journal of Finance*, 57(3), 1047–1091.
- Amengual, D., 2008, “The term structure of variance risk premia,” Working Paper.
- Amengual, D., and D. Xiu, 2015, “Resolution of Policy Uncertainty and Sudden Declines in Volatility,” Chicago Booth Paper No. 13-78.
- Andersen, T., N. Fusari, and V. Todorov, 2015, “The Risk Premia Embedded in Index Options,” *Journal of Financial Economics*, Forthcoming.
- Andersen, T. G., L. Benzoni, and J. Lund, 2002, “An Empirical Investigation of Continuous-Time Equity Return Models,” *Journal of Finance*, 57(3), 1239–1284.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, “The Distribution of Realized Stock Return Volatility,” *Journal of Financial Economics*, 61, 43–76.
- Andries, M., T. Eisenbach, M. Schmalz, and Y. Wang, 2015, “The Term Structure of the Price of Variance Risk,” Federal Reserve Bank of New York Staff Report No. 736.
- Bakshi, G., N. Kapadia, and D. Madan, 2003, “Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options,” *Review of Financial Studies*, 16(1), 101–143.
- Bali, T. G., and H. Zhou, 2016, “Risk, Uncertainty and Expected Returns,” *Journal of Financial and Quantitative Analysis*, 51(3), 707–735.

- Bates, D. S., 1996, “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options,” *Review of Financial Studies*, 9(1), 69–107.
- , 2000a, “Post-87 Crash Fears in the S&P 500 Futures Option Market,” *Journal of Econometrics*, 94(1-2), 181–238.
- , 2000b, “Post-87 Crash Fears in the S&P 500 Futures Options,” *Journal of Econometrics*, 94(1-2), 181–238.
- , 2012, “U.S. stock market crash risk, 1926-2010,” *Journal of Financial Economics*, 105(2), 229 – 259.
- Bayer, C., J. Gatheral, and M. Karlsmark, 2013, “Fast Ninomiya-Victoir calibration of the Double-Mean-Reverting Model,” *Quantitative Finance*, 13(11), 1813–1829.
- Bekaert, G., and M. Hoerova, 2014, “The VIX, the variance premium and stock market volatility,” *Journal of Econometrics*, 183, 181–192.
- Bollerslev, T., G. Tauchen, and H. Zhou, 2009, “Expected stock returns and variance risk premia,” *Review of Financial Studies*, 22, 4463–92.
- Bollerslev, T., and V. Todorov, 2011, “Tails, Fears and Risk Premia,” *Journal of Finance*, 66(6), 2165–2211.
- Breedon, D. T., and R. H. Litzenberger, 1978, “Prices of State-Contingent Claims Implicit in Option Prices,” *Journal of Business*, 51(4), 621–651.
- Broadie, M., M. Chernov, and M. Johannes, 2007, “Model Specification and Risk Premia: Evidence from Futures Options,” *Journal of Finance*, 62(3), 1453–1490.
- Carr, P., and D. B. Madan, 1999, “Option valuation using the fast Fourier transform,” *Journal of Computational Finance*, 2(4), 1–18.
- Chernov, M., A. R. Gallant, E. Ghysels, and G. T. Tauchen, 2003, “Alternative Models of Stock Price Dynamics,” *Journal of Econometrics*, 116(1-2), 225 – 257.
- Christoffersen, P., S. Heston, and K. Jacobs, 2009, “The Shape and Term Structure of the Index

- Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well,” *Management Science*, 55(12), 1914–1932.
- Christoffersen, P., K. Jacobs, and K. Mimouni, 2010, “Models for S&P 500 Dynamics: Evidence from Realized Volatility, Daily Returns and Options Prices,” *Review of Financial Studies*, 23(8), 3141–3189.
- Chung, S.-L., W.-C. Tsai, Y.-H. Wang, and P.-S. Weng, 2011, “The Information Content of the S&P 500 Index and VIX Options on the Dynamics of the S&P 500 Index,” *The Journal of Futures Markets*, 31(12), 1170–1201.
- Cont, R., and T. Kokholm, 2013, “A Consistent Pricing Model for Index Options and Volatility Derivatives,” *Mathematical Finance*, 23(2), 248–274.
- Dew-Becker, I., S. Giglio, A. Le, and M. Rodriguez, 2015, “The price of variance risk,” .
- Drechsler, I., and A. Yaron, 2011, “What’s Vol Got to Do with It?,” *Review of Financial Studies*, 24, 1–45.
- Duan, J.-C., and C.-Y. Yeh, 2010, “Jump and volatility risk premiums implied by VIX,” *Journal of Economic Dynamics and Control*, 34(11), 2232–2244.
- , 2011, “Price and Volatility Dynamics Implied by the VIX Term Structure,” working paper, NUS RMI Working Paper No. 11/05.
- Duffie, D., J. Pan, and K. J. Singleton, 2000, “Transform Analysis and Asset Pricing for Affine Jump-Diffusions,” *Econometrica*, 68(6), 1343–1376.
- Durham, G. B., 2013, “Risk-neutral Modeling with Affine and Nonaffine Models,” *Journal of Financial Econometrics*, 11(4), 650–681.
- Egloff, D., M. Leippold, and L. Wu, 2010, “Valuation and Optimal Investing in Variance Swaps,” *Journal of Financial and Quantitative Analysis*, 45(5), 1279–1310.
- Eraker, B., 2004, “Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices,” *Journal of Finance*, 59(3), 1367–1404.

- Fang, F., and C. W. Oosterlee, 2008, “A novel pricing method for European options based on Fourier-cosine series expansions,” *SIAM Journal on Scientific Computing*, 31(2), 826–848.
- Ferriani, F., and S. Pastorello, 2012, “Estimating and Testing Non-Affine Option Pricing Models With a Large Unbalanced Panel of Options,” *The Econometrics Journal*, 15(2), 171–203.
- Filipović, D., E. Gourier, and L. Mancini, 2016, “Quadratic variance swap models,” *Journal of Financial Economics*, 119(1), 44–68.
- Gatheral, J., 2008, “Consistent Modeling of SPX and VIX options,” The Fifth World Congress of the Bachelier Finance Society, London, July 2008.
- Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation,” *Radar and Signal Processing, IEE Proceedings F*, 140(2), 107–113.
- Gruber, P. H., C. Tebaldi, and F. Trojani, 2015, “The Price of the Smile and Variance Risk Premia,” Swiss Finance Institute Research Paper No. 15-36.
- Hansen, N., and A. Ostermeier, 1996, “Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation,” *Proceedings of the 1996 IEEE Conference on Evolutionary Computation (ICEC 96)*, pp. 312–317.
- Hu, F., and J. V. Zidek, 2002, “The Weighted Likelihood,” *The Canadian Journal of Statistics*, 30(3), 347–371.
- Huang, D., and I. Shaliastovich, 2015, “Volatility-of-Volatility Risk,” .
- Jacod, J., and V. Todorov, 2010, “Do Price and Volatility Jump Together?,” *Annals of Applied Probability*, 20(4), 1425–1469.
- Jiang, G. J., and Y. S. Tian, 2007, “Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index,” *The Journal of Derivatives*, 14(3), 35–60.
- Johannes, M. S., N. G. Polson, and J. R. Stroud, 2009, “Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices,” *Review of Financial Studies*, 22(7), 2759–2799.

- Johnson, T. L., 2012, “Equity Risk Premia and the VIX Term Structure,” Working paper, Stanford University.
- Jones, C. S., 2003, “The Dynamics of Stochastic Volatility: Evidence From Underlying and Options Markets,” *Journal of Econometrics*, 116(1-2), 181–224.
- Kaeck, A., and C. Alexander, 2012, “Volatility dynamics for the S&P 500: Further evidence from non-affine, multi-factor jump diffusions,” *Journal of Banking & Finance*, 36(11), 3110 – 3121.
- Kelly, B., and H. Jiang, 2014, “Tail risk and asset prices,” *Review of Financial Studies*, 27, 2841–2871.
- Kloeden, P. E., and E. Platen, 1992, *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, Germany.
- Lindström, E., J. Ströjby, M. Brodén, M. Wiktorsson, and J. Holst, 2008, “Sequential calibration of options,” *Computational Statistics & Data Analysis*, 52(6), 2877–2891.
- Mencía, J., and E. Sentana, 2013, “Valuation of VIX derivatives,” *Journal of Financial Economics*, 108(2), 367 – 391.
- Ornthanalai, C., 2014, “Lévy Jump Risk: Evidence from Options and Returns,” *Journal of Financial Economics*, 112, 69–90.
- Pan, J., 2002, “The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study,” *Journal of Financial Economics*, 63(1), 3–50.
- Papanicolaou, A., and R. Sircar, 2013, “A regime-switching Heston model for VIX and S&P 500 implied volatilities,” *Quantitative Finance*, Forthcoming.
- Pitt, M. K., 2002, “Smooth particle filters for likelihood evaluation and maximisation,” working paper 651, The Warwick Economics Research Paper Series.
- Pitt, M. K., and N. Shephard, 1999, “Filtering via Simulation: Auxiliary Particle Filters,” *Journal of the American Statistical Association*, 94(446), 590–599.
- Rebonato, R., and T. Cardoso, 2004, “Unconstrained Fitting of Implied Volatility Surfaces Using a Mixture of Normals,” *Journal of Risk*, 7(1), 55–74.

- Sepp, A., 2008a, “Pricing Options on Realized Variance in the Heston Model with Jumps in Returns and Volatility,” *Journal of Computational Finance*, 11(4), 33–70.
- , 2008b, “VIX Option Pricing in a Jump-Diffusion Model,” *Risk Magazine*, pp. 84–89.
- Song, Z., and D. Xiu, 2015, “A Tale of Two Option Markets: Pricing Kernels and Volatility Risk,” *Journal of Econometrics*, Forthcoming.
- Storn, R., 1996, “On the usage of differential evolution for function optimization,” in *Biennial Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, pp. 519–523.
- Todorov, V., 2010, “Variance Risk Premium Dynamics: The Role of Jumps,” *Review of Financial Studies*, 23(1), 345–383.
- Todorov, V., and G. Tauchen, 2011, “Volatility Jumps,” *Journal of Business and Economic Statistics*, 29(3), 356–371.
- van Binsbergen, J. H., W. Hueskes, R. S. J. Koijen, and E. B. Vrugt, 2013, “Equity Yields,” *Journal of Financial Economics*, 110(3), 503–519.
- Wu, L., 2011, “Variance Dynamics: Joint Evidence from Options and High-Frequency Returns,” *Journal of Econometrics*, 160(1), 280–287.
- Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2005, “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data,” *Journal of the American Statistical Association*, 100, 1394–1411.

Appendix

A Affine dependence of the VIX² on v_t , m_t and u_t

The expressions for the coefficients α_{VIX^2} , β_{VIX^2} , γ_{VIX^2} and δ_{VIX^2} in Proposition 2.1 are given by

$$\alpha_{\text{VIX}^2} = \left(1 + 2\lambda_1^{(+)}C^{(+)} + 2\lambda_1^{(-)}C^{(-)}\right) A(\tau_{\text{VIX}}), \quad (\text{A.1})$$

$$\beta_{\text{VIX}^2} = \left(1 + 2\lambda_1^{(+)}C^{(+)} + 2\lambda_1^{(-)}C^{(-)}\right) B(\tau_{\text{VIX}}) + \left(2\lambda_2^{(+)}C^{(+)} + 2\lambda_2^{(-)}C^{(-)}\right) \hat{A}(\tau_{\text{VIX}}), \quad (\text{A.2})$$

$$\gamma_{\text{VIX}^2} = \left(1 + 2\lambda_1^{(+)}C^{(+)} + 2\lambda_1^{(-)}C^{(-)}\right) D(\tau_{\text{VIX}}) + \left(2\lambda_3^{(+)}C^{(+)} + 2\lambda_3^{(-)}C^{(-)}\right) \bar{A}(\tau_{\text{VIX}}), \quad (\text{A.3})$$

$$\delta_{\text{VIX}^2} = 2(\lambda_0^{(+)}C^{(+)} + \lambda_0^{(-)}C^{(-)}) + \left(1 + 2\lambda_1^{(+)}C^{(+)} + 2\lambda_1^{(-)}C^{(-)}\right) G + \quad (\text{A.4})$$

$$\left(2\lambda_2^{(+)}C^{(+)} + 2\lambda_2^{(-)}C^{(-)}\right) \hat{B}(\tau_{\text{VIX}}) + \left(2\lambda_3^{(+)}C^{(+)} + 2\lambda_3^{(-)}C^{(-)}\right) \bar{B}(\tau_{\text{VIX}}), \quad (\text{A.5})$$

where $C^{(+/-)} := \frac{1}{1 - \mu_V^{(+/-)}}$, τ_{VIX} is 30 days and the remaining coefficients are available in closed form from:

$$A(\tau) = \frac{1}{\tau_{\text{VIX}}} \int_t^{t+\tau_{\text{VIX}}} \alpha_v(t, s) ds; \quad B(\tau) = \frac{1}{\tau_{\text{VIX}}} \int_t^{t+\tau_{\text{VIX}}} \beta_v(t, s) ds; \quad (\text{A.6})$$

$$D(\tau) = \frac{1}{\tau_{\text{VIX}}} \int_t^{t+\tau_{\text{VIX}}} \gamma_v(t, s) ds; \quad G(\tau) = \frac{1}{\tau_{\text{VIX}}} \int_t^{t+\tau_{\text{VIX}}} \delta_v(t, s) ds; \quad (\text{A.7})$$

$$\hat{A}(\tau) = \frac{1}{\tau_{\text{VIX}}} \int_t^{t+\tau_{\text{VIX}}} \alpha_m(t, s) ds; \quad \hat{B}(\tau) = \frac{1}{\tau_{\text{VIX}}} \int_t^{t+\tau_{\text{VIX}}} \beta_m(t, s) ds; \quad (\text{A.8})$$

$$\bar{A}(\tau) = \frac{1}{\tau_{\text{VIX}}} \int_t^{t+\tau_{\text{VIX}}} \alpha_u(t, s) ds; \quad \bar{B}(\tau) = \frac{1}{\tau_{\text{VIX}}} \int_t^{t+\tau_{\text{VIX}}} \beta_u(t, s) ds; \quad (\text{A.9})$$

with the functions α_v , β_v , γ_v , δ_v , α_m , β_m , α_u and β_u such that, for $0 \leq t < s$:

$$\mathbb{E}_t^{\mathbb{Q}}[v_s] = \alpha_v(t, s)v_t + \beta_v(t, s)m_t + \gamma_v(t, s)u_t + \delta_v(t, s); \quad (\text{A.10})$$

$$\mathbb{E}_t^{\mathbb{Q}}[m_s] = \alpha_m(t, s)m_t + \beta_m(t, s); \quad (\text{A.11})$$

$$\mathbb{E}_t^{\mathbb{Q}}[u_s] = \alpha_u(t, s)u_t + \beta_u(t, s). \quad (\text{A.12})$$

B Characteristic functions

Assuming that the characteristic functions of the processes Y , VIX^2 take the exponential affine form

$$\begin{aligned}\Psi_{VIX_T^2}(t, v, m, u; \omega) &= \mathbb{E}_t^{\mathbb{Q}} \left[e^{\omega VIX_T^2} \right] = e^{\alpha(T-t) + \beta(T-t)v + \gamma(T-t)m + \delta(T-t)u}, \\ \Psi_{Y_T}(t, v, m, u; \omega) &= \mathbb{E}_t^{\mathbb{Q}} \left[e^{\omega Y_T} \right] = e^{\alpha_Y(T-t) + \beta_Y(T-t)y + \gamma_Y(T-t)v + \delta_Y(T-t)m + \xi(T-t)u},\end{aligned}$$

with $\omega \in \mathbb{C}$, the coefficients in the definition of $\Psi_{VIX_T^2}$ satisfy the following ODEs, with $\tau = T - t$:¹⁶

$$\begin{aligned}-\alpha'(\tau) + \gamma(\tau)\kappa_m\theta_m + \delta(\tau)\kappa_u\theta_u + \lambda_0^{(-)} \left(\frac{1}{1 - \beta(\tau)\nu_v^{(+)}} - 1 \right) + \lambda_0^{(+)} \left(\frac{1}{1 - \beta(\tau)\nu_v^{(-)}} - 1 \right) &= 0 \\ -\beta'(\tau) - \beta(\tau)\kappa_v + \frac{1}{2}\sigma_v^2\beta^2(\tau) + \lambda_1^{(-)} \left(\frac{1}{1 - \beta(\tau)\nu_v^{(+)}} - 1 \right) + \lambda_1^{(+)} \left(\frac{1}{1 - \beta(\tau)\nu_v^{(-)}} - 1 \right) &= 0 \\ -\gamma'(\tau) - \gamma(\tau)\kappa_m + \frac{1}{2}\sigma_m^2\gamma^2(\tau) + \kappa_v\beta(\tau) + \lambda_2^{(-)} \left(\frac{1}{1 - \beta(\tau)\nu_v^{(+)}} - 1 \right) + \lambda_2^{(+)} \left(\frac{1}{1 - \beta(\tau)\nu_v^{(-)}} - 1 \right) &= 0 \\ -\delta'(\tau) - \delta(\tau)\kappa_u + \frac{1}{2}\delta^2(\tau)\sigma_u^2 + \lambda_3^{(-)} \left(\frac{1}{1 - \beta(\tau)\nu_v^{(+)}} - 1 \right) + \lambda_3^{(+)} \left(\frac{1}{1 - \beta(\tau)\nu_v^{(-)}} - 1 \right) &= 0\end{aligned}$$

$\forall t \in (0, T]$, with boundary conditions $\alpha(0) = \omega\gamma_{VIX^2}$, $\beta(0) = \omega\alpha_{VIX^2}$, $\gamma(0) = \omega\beta_{VIX^2}$ and $\delta(0) = \omega\gamma_{VIX^2}$, where the coefficients α_{VIX^2} , β_{VIX^2} , γ_{VIX^2} and δ_{VIX^2} are defined in Appendix A.

¹⁶This relies on the fact that the Poisson processes driving the positive and negative jumps in Y_t are independent.

The coefficients of Ψ_{Y_T} satisfy the following ODEs for $t \in (0, T]$:

$$\begin{aligned}
& -\alpha'_Y(\tau) + \beta_Y(\tau) \left(-\lambda_0^{(+)} \left(\frac{1}{1 - \mu_Y^{(+)}} - 1 \right) - \lambda_0^{(-)} \left(\frac{1}{1 - \mu_Y^{(-)}} - 1 \right) \right) + \delta_Y(T - t) \kappa_m \theta_m + \xi_Y(\tau) \kappa_u \theta_u + \\
& \quad \lambda_0^{(+)} \left(\frac{1}{(1 - \beta_Y(\tau) \mu_Y^{(+)}) (1 - \gamma_Y(\tau) \nu_v^{(-)})} - 1 \right) + \lambda_0^{(-)} \left(\frac{1}{(1 - \beta_Y(\tau) \mu_Y^{(-)}) (1 - \gamma_Y(\tau) \nu_v^{(+)})} - 1 \right) = 0 \\
& -\beta'_Y(\tau) = 0 \\
& -\gamma'_Y(\tau) + \beta_Y(\tau) \left(-\lambda_1^{(+)} \left(\frac{1}{1 - \mu_Y^{(+)}} - 1 \right) - \lambda_1^{(-)} \left(\frac{1}{1 - \mu_Y^{(-)}} - 1 \right) - \frac{1}{2} \right) - \gamma_Y(\tau) \kappa_v + \frac{1}{2} \beta_Y(\tau)^2 + \\
& \quad \frac{1}{2} \gamma_Y(\tau)^2 \sigma_v^2 + \beta_Y(\tau) \gamma_Y(\tau) \sigma_v \rho_{Y,v} + \lambda_1^{(+)} \left(\frac{1}{(1 - \beta_Y(\tau) \mu_Y^{(+)}) (1 - \gamma_Y(\tau) \nu_v^{(-)})} - 1 \right) + \\
& \quad \lambda_1^{(-)} \left(\frac{1}{(1 - \beta_Y(\tau) \mu_Y^{(-)}) (1 - \gamma_Y(\tau) \nu_v^{(+)})} - 1 \right) = 0 \\
& -\delta'_Y(\tau) + \beta_Y(\tau) \left(-\lambda_2^{(+)} \left(\frac{1}{1 - \mu_Y^{(+)}} - 1 \right) - \lambda_2^{(-)} \left(\frac{1}{1 - \mu_Y^{(-)}} - 1 \right) \right) + \gamma_Y(\tau) \kappa_v - \delta_Y(\tau) \kappa_m + \frac{1}{2} \delta_Y(\tau)^2 \sigma_m^2 + \\
& \quad \lambda_2^{(+)} \left(\frac{1}{(1 - \beta_Y(\tau) \mu_Y^{(+)}) (1 - \gamma_Y(\tau) \nu_v^{(-)})} - 1 \right) + \lambda_2^{(-)} \left(\frac{1}{(1 - \beta_Y(\tau) \mu_Y^{(-)}) (1 - \gamma_Y(\tau) \nu_v^{(+)})} - 1 \right) = 0 \\
& -\xi'_Y(\tau) + \beta_Y(\tau) \left(-\lambda_3^{(+)} \left(\frac{1}{1 - \mu_Y^{(+)}} - 1 \right) - \lambda_3^{(-)} \left(\frac{1}{1 - \mu_Y^{(-)}} - 1 \right) \right) - \xi_Y(\tau) \kappa_u + \frac{1}{2} \xi_Y(T - t)^2 \sigma_u^2 + \\
& \quad \lambda_3^{(+)} \left(\frac{1}{(1 - \beta_Y(\tau) \mu_Y^{(+)}) (1 - \gamma_Y(\tau) \nu_v^{(-)})} - 1 \right) + \lambda_3^{(-)} \left(\frac{1}{(1 - \beta_Y(\tau) \mu_Y^{(-)}) (1 - \gamma_Y(\tau) \nu_v^{(+)})} - 1 \right) = 0,
\end{aligned}$$

with boundary conditions $\alpha_Y(0) = 0$, $\beta_Y(0) = \omega$, $\gamma_Y(0) = 0$, $\delta_Y(0) = 0$ and $\xi_Y(0) = 0$.

In practice, the ODEs are solved numerically.

C Coefficients for the Fourier cosine expansion

Here we give the expression for $U_n^{\text{VIX}^2}$, the Fourier cosine transform of the VIX options' payoff. To ease notation, we drop the subscript VIX for $a_{\text{VIX}}, b_{\text{VIX}}$ and define $\omega_n := \frac{n\pi}{b-a}$. For $n > 0$, we obtain

$$\begin{aligned} U_n^{\text{VIX}^2} &= \int_a^b (\sqrt{x} - K)^+ \cos(\omega_n(x - a)) dx \\ &= \frac{2}{b-a} \text{Re} \left\{ e^{-i\omega_n a} \left[\frac{\sqrt{b} e^{-i\omega_n b}}{i\omega_n} + \frac{\sqrt{\pi}}{2(-i\omega_n)^{3/2}} \left(\text{erfz}(\sqrt{-i\omega_n b}) - \text{erfz}(K\sqrt{-i\omega_n}) \right) \right] \right\}. \end{aligned} \quad (\text{C.1})$$

where erfz is the error function for complex numbers.

For $n = 0$,

$$U_0^{\text{VIX}^2} = \frac{2}{b-a} \left[\frac{2}{3} b^{3/2} - Kb + \frac{1}{3} K^3 \right]. \quad (\text{C.2})$$

D Joint calibration

Before we bring our models to the time series of data, we carry out a joint calibration exercise using the cross-section of S&P 500 and VIX options on a particular date. This exercise gives us some guidance for model design and allows us to reduce the set of models to be estimated on a time series of options' data. If a model is not flexible enough to jointly reproduce the implied volatility patterns of both markets on a single date, the \mathbb{Q} -dynamics of the model is not sufficiently rich to accurately price both the S&P 500 and VIX derivatives jointly, and we can safely discard this model from further consideration.

Let us fix a date t and consider $\{IV_{\text{SPX},i}^{Mkt}\}_{i=1 \dots N_{\text{SPX}}}$, the set of N_{SPX} market implied volatilities of S&P 500 options for strikes $\{K_i\}$ and maturities $\{T_i\}$. We denote by $\{IV_{\text{VIX},j}^{Mkt}\}_{j=1 \dots N_{\text{VIX}}}$ the set of N_{VIX} market implied volatilities of VIX options. To estimate the parameters, we minimize the root

mean squared error (RMSE) between the market and model implied volatilities:¹⁷

$$\text{RMSE}_{\mathcal{M}}(t) := \sqrt{\frac{1}{N_{\mathcal{M}}} \sum_{1 \leq i \leq I} \left(IV_{\mathcal{M},i}^{Mkt} - IV_{\mathcal{M},i}^{Mod} \right)^2}, \quad \mathcal{M} \in \{\text{SPX}, \text{VIX}\}, \quad (\text{D.1})$$

$$\text{RMSE}(t) := \frac{1}{2} (\text{RMSE}_{\text{SPX}}(t) + \text{RMSE}_{\text{VIX}}(t)). \quad (\text{D.2})$$

We use two global optimizers to cope with the non-convexity of the calibration problem and the potential existence of multiple local minima, 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).

For our calibration, we choose a date on which the markets were under stress, namely May 5, 2010 at the beginning of the European sovereign debt crisis. After cleaning our data as described previously, we have 91 VIX options at six different maturities (from 0.04 to 0.46 years) and 486 S&P 500 options at eleven different maturities (from 0.05 to 0.91 years) available. We emphasize that we perform a joint calibration. Hence, all this data is entered as input to minimize the total RMSE in (D.2) from the VIX and the S&P 500 market simultaneously across all available maturities and moneyness.

In Figure 9, we plot the market and model implied volatilities for the S&P 500 (Panels A, C, E) and the VIX (Panels B, D, F) for two maturity slices each. For the S&P 500 options, we choose the two maturities $T = 0.05$ and $T = 0.3$, and for the VIX options, $T = 0.04$ and $T = 0.36$. As candidate models, we choose different sub-specifications of the general model (SVJ3) presented in Equations (1)–(7), namely the Heston model, the Heston model with negative jumps in returns and positive jumps in volatility (SVJ), and the latter model augmented with the factor m (SVJ2). Richer model are found to not provide any improvements on one day of data, and are examined in the time-series consistent estimation.

[Figure 9 about here.]

From Panel A, Figure 9, we observe that the Heston model provides reasonable results for the S&P

¹⁷Alternatively, we checked that using distances taking into account the bid–ask spread of IVs as in Cont and Kokholm (2013) does not significantly change the quality of fit. Instead of the RMSE, we also looked at average relative errors (ARE). However, this does not affect our conclusions. The results using ARE are available upon request.

500 market. However, for the VIX market (Panel B), the Heston model clearly fails to reproduce one of the stylized facts of VIX option markets: the positive skew of the implied volatility surface. This failure is most pronounced for the short-term options, where the Heston model generates a significant negative skew. The results for the SVJ model look much more promising. Just by adding jump components to the returns and volatility process, we can now generate the positive skew in the VIX market (Panel D), while providing an almost perfect fit for the S&P 500 options market. The SVJ model only struggles at the short end of the VIX implied volatility surface. This shortcoming disappears when we extend the SVJ specification to the SVJ2 model by adding the factor m . Doing so gives us not only a remarkable fit for the S&P 500 but also for the VIX options market (Panel F). Looking at the RMSEs of the SVJ and SVJ2 models, we find that the SVJ provides an RMSE_{SPX} of 1.27% and an RMSE_{VIX} of 11.60%. The SVJ2 model yields 1.17% and 5.15%, respectively. Hence, while the two models are comparable in terms of their performance on the S&P 500 options market, there is an obvious difference in the VIX market on the chosen date.

In unreported results, we perform calibration exercises on other days, also including calm periods. Irrespective of the day, we observe that the SVJ and SVJ2 models perform comparably on the S&P 500 options market, both fitting the data very well. In contrast, we find 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.¹⁸ Therefore, we conclude from our calibration exercise that we can discard the Heston model from further analysis and that jumps in the volatility are necessary.

E Particle filter

The log-likelihood of a time series of $n + 1$ observations with joint density p , conditional on a set of parameters Θ , is equal to

$$\log p(y^{t_n} | \Theta) = \log p(y_{t_0}, \dots, y_{t_n} | \Theta) = \sum_{k=1}^n \log p(y_{t_k} | y^{t_{k-1}}, \Theta) + \log p(y_{t_0} | \Theta) \quad (\text{E.1})$$

¹⁸Our findings are consistent with Gatheral (2008), who shows that the Heston model is incapable of reproducing the positive skew in VIX IVs, and with Sepp (2008a,b), who finds that incorporating positive jumps in the volatility dynamics into the Heston model removes this shortcoming.

where, by the Law of Total Probability,

$$p(y_{t_k}|y^{t_{k-1}}, \Theta) = \int p(y_{t_k}|L_{t_k}, \Theta)p(L_{t_k}|y^{t_{k-1}}, \Theta)dL_{t_k}. \quad (\text{E.2})$$

Given an initial density $p(L_{t_0}|\Theta)$, the transition density of the state variables $p(L_{t_k}|L_{t_{k-1}}, \Theta)$ and the likelihood function $p(y_{t_k}|L_{t_k}, \Theta)$, filtering methods allow us to estimate the distribution $p(L_{t_k}|y^{t_k}, \Theta)$ of the current state at time $t_k = k\Delta t$, given all observations up to that time. In the following, we simplify notation and drop the subscript for the conditioning on the parameters Θ . The filtering density is given by Bayes's formula,

$$p(L_{t_k}|y^{t_k}) \propto p(y_{t_k}|L_{t_k})p(L_{t_k}|y^{t_{k-1}}). \quad (\text{E.3})$$

The likelihood function is known, but the predictive distribution of the state is not. It is given by the following integral, which involves the previous filtering density.

$$p(L_{t_k}|y^{t_{k-1}}) = \int p(L_{t_k}|L_{t_{k-1}})p(L_{t_{k-1}}|y^{t_{k-1}})dL_{t_{k-1}} \quad (\text{E.4})$$

The key idea is to approximate the posterior density function of the latent variables $p(L_{t_k}|y^{t_k})$ by a sum of point masses positioned at strategic points, called particles, $\{L_{t_k}^{(i)}\}_{1 \leq i \leq n_p}$:

$$\hat{p}(L_{t_k}|y^{t_k}) = \sum_{i=1}^{n_p} \pi_{t_k}^{(i)} \delta(L_{t_k} - L_{t_k}^{(i)}), \quad (\text{E.5})$$

where $\pi_{t_k}^{(i)}$ denotes the normalized importance weight for particle i , $\delta(\cdot)$ is the Dirac delta function, and n_p is the number of support points (particles) for $\hat{p}(L_{t_k}|y^{t_k})$. Then, we can recursively calculate the filtering density by

$$\begin{aligned} \hat{p}(L_{t_k}|y^{t_k}) &\propto \int p(y_{t_k}|L_{t_k})p(L_{t_k}|L_{t_{k-1}})\hat{p}(L_{t_{k-1}}|y^{t_{k-1}})dL_{t_{k-1}} \\ &= \sum_{i=1}^{n_p} p(y_{t_k}|L_{t_k})p(L_{t_k}|L_{t_{k-1}}^{(i)})\pi_{t_{k-1}}^{(i)}. \end{aligned} \quad (\text{E.6})$$

To implement the particle filter, one needs to be able to simulate at every time t_k a number n_p of particles $L_{t_k}^{(i)}, i = 1, \dots, n_p$ from $p(L_{t_k}|y^{t_k-1})$ and to be able to evaluate $p(y_{t_k}|L_{t_k}^{(i)})$. Based on these simulated particles, we can approximate $p(y_{t_k}|y^{t_k-1})$ by

$$p(y_{t_k}|y^{t_k-1}) \approx \frac{1}{n_p} \sum_{i=1}^{n_p} p(y_{t_k}|L_{t_k}^{(i)}). \quad (\text{E.7})$$

The available observations are the S&P 500 daily returns, the VIX levels, and the option prices on both indexes. In the following, the notation is as in Section 4.2. We used $n_p = 15,000$ particles on days when the observations contain option prices and $n_p = 8,000$ when the observations are only composed of the S&P500 returns and VIX index levels. Larger numbers of particles did not change our estimates, but increased the computational burden.

The filtering algorithm can be decomposed into the following steps.

Step 1: Initialization. We simulate n_p initial particles for the latent variables $\{v_{t_0}^{(i)}, m_{t_0}^{(i)}, U_{t_0}^{(i)}\}_{i=1, \dots, n_p}$ which are compatible with the initial value of the VIX squared, i.e., given the specification (20). The following steps are repeated for each time step t_k in the grid from $k = 0$ to $k = M - 1$.

Step 2: First-stage resampling. At this point, we assume that we have n_p particles (i.e., possible values of m_t and v_t) at time t_k given all observations y^{t_k} up to t_k . At time t_{k+1} , there are new observations y_{k+1} . The goal of this step is to retain, 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 observations y_{k+1} . For this purpose, we assign a weight (the so called “first-stage weights”) to each particle, which is proportional to the likelihood of new market observations $y_{t_{k+1}}$ 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 indexes.

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 mass function. Conditioning on the number of jumps in ΔY_t (or equivalently in Δv_t) and Δm_t gives

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

Given that we use daily observations, we limit the possible number of jumps of the Poisson random variables $\Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)}$ to zero or one (this Bernoulli approximation is found to be very accurate in Johannes, Polson, and Stroud (2009)). We recall that the new observation is composed of the S&P 500 returns and the VIX level $y_{t_{k+1}} = (\Delta Y_{t_k} = Y_{t_{k+1}} - Y_{t_k}, \text{VIX}_{t_{k+1}}^2)$. Since the $\text{VIX}_{t_{k+1}}^2$ is a sum of normal distributions and no more than 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 (a generalization of a Bernoulli distribution) which is supported in a certain number of (the corresponding exponential distribution's) quantiles.¹⁹ 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 $y_{t_{k+1}}$, we resample (with replacement) the particles according to a stratified resampling scheme:²⁰

$$z^{(i)} \sim \text{StratRes}(n_p, \omega_{t_{k+1}}^{(1)}, \dots, \omega_{t_{k+1}}^{(n_p)}).$$

This makes it possible to create a new sample of n_p latent factors $\{L_{t_k}^{(j)}\}_{1 \leq j \leq n_p}$ 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 $L_{t_k}^{(i)}$ and the new observations $y_{t_{k+1}}$:

$$L_{t_{k+1}}^{(i)} \sim p(L_{t_{k+1}}|L_{t_k}^{(i)}, y_{t_{k+1}}).$$

¹⁹Robustness tests were performed on simulated data to check that the choice of quantiles was appropriate.

²⁰We checked that using a multinomial or stratified resampling scheme gives similar results.

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

Step 3: Generating the jumps. We calculate the joint probability of jumps in ΔY_t (or equivalently in Δv_t) between t_k and t_{k+1} using

$$\mathbb{P}(\Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)} | y_{t_{k+1}}) \propto p(y_{t_{k+1}} | \Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)}) \mathbb{P}(\Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)}). \quad (\text{E.8})$$

Conditionally on the jump sizes, the first part of the right-hand side has already been calculated in the first-stage weights. Using Bayes's rule, we get an approximation for $\mathbb{P}(\Delta N_{t_k}^{(+)}, \Delta N_{t_k}^{(-)} | y_{t_{k+1}})$.

We simulate the jump sizes for Δv_t and Δm_t according to their exponential law.

Step 4: Propagating the latent factors. The latent factors v , m and u are propagated following a Milstein discretization scheme of the SDE. See Kloeden and Platen (1992) for details. We use the full truncation method to prevent them from taking negative values.

Step 5: Computing the filtering density. At this point, the newly generated particles $\{L_{t_{k+1}}^{(i)}\}_{1 \leq i \leq n_p}$ are a sample of $p(L_{t_{k+1}} | y^{t_{k+1}})$. We now calculate the *second-stage weights* $\{\pi_{t_{k+1}}^{(i)}\}_{1 \leq i \leq n_p}$ which approximate the probabilities $p(L_{t_{k+1}}^{(i)} | y^{t_{k+1}})$ 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 $L_{t_{k+1}}^{(i)}$, with a correction related to the proposal density

$$\pi_{t_{k+1}}^{(i)} \propto \frac{p(L_{t_{k+1}}^{(i)} | L_{t_k}^{(i)}) p(y_{t_{k+1}} | L_{t_{k+1}}^{(i)})}{\omega_{t_{k+1}}^{z^{(i)}} q(L_{t_{k+1}}^{(i)} | L_{t_k}^{(i)}, y_{t_{k+1}})}.$$

The posterior distribution of the state variables is approximated by

$$\hat{p}(L_{t_{k+1}} | y^{t_{k+1}}) = \sum_{i=1}^{n_p} \pi_{t_{k+1}}^{(i)} \delta(L_{t_{k+1}} - L_{t_{k+1}}^{(i)}).$$

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

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 the particles is then estimated by the average of the second-stage weights over the particles

$$\hat{p}(y^{t_n}|\Theta, \mathcal{M}) = \prod_{k=1}^M \hat{p}(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M})\hat{p}(y_{t_0}|\Theta, \mathcal{M}),$$

where $\hat{p}(y_{t_k}|y^{t_{k-1}}, \Theta, \mathcal{M}) = \frac{1}{n_p} \sum_{i=1}^{n_p} \pi_{t_k}^{(i)}$.

Finally, we use the Weighted Likelihood function analyzed in Hu and Zidek (2002), in order to assign comparable weights to the different datasets, and to ensure that the estimation is not dominated by S&P 500 options. Such procedure has been used by Ornathanalai (2014) to estimate a model with Lévy jumps to S&P 500 options and returns.

F Specific data treatment for the particle filter

The S&P 500 options dataset contains a large amount of ATM options compared to OTM and deep OTM options. If we use the filter (within the maximum likelihood procedure) on this entire dataset, the fitting of ATM options will be its priority rather than (deep) OTM options. Given the formula of 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 the extreme events contained in the data to be incorporated into the models. For this reason, we interpolate the S&P 500 IV slices and re-sample the option prices from the resulting parametric fit uniformly with respect to moneynesses.²¹ Other advantages of our use of interpolation is that the resulting data is arbitrage free, we have fewer points for each slice (but still accurately representing the information of each

²¹It 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.

slice), thus reducing the computational complexity.²²

For the interpolation, we use 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 RMSE of the S&P 500 implied volatilities parametric fits are on average around 0.25% and we therefore do not lose information, especially given the market bid–ask spread. Finally, using the parametric fit, we can sample a fixed number (we have chosen 15) of “market option prices” for the desired strikes. We have chosen to resample the option prices from each parametric slice uniformly in the strike (or, equivalently, the moneyness). We however do not resample the 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.

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). Therefore all available VIX option prices are used.

As the datasets comprise a large number of options (up to 600 a day), it is unfeasible to calculate the option prices every day for every particle. As a consequence, we follow Pan (2002) and Johannes, Polson, and Stroud (2009), among others, and use weekly (Wednesday) options data. Furthermore, this eliminates beginning-of-week and end-of-week effects. Our particle filter uses daily time steps and incorporates information on the underlying indexes on a daily basis (i.e., only options are considered weekly).

We decompose the time series of observations into two periods. The first period is from March 1st, 2006 to Feb 28, 2009 (shortly after the VIX index increased to its highest point). This was a rather calm period,²³ that we will use as the in-sample estimation period. Our out-of-sample period starts on March 1st, 2009 and ends on October 29, 2010. This period includes very high levels of volatility (for implied volatilities from the S&P 500 and VIX options as well as the VIX index

²²Since we have considered mid-prices and because of synchronization issues between the underlying and the options, implied volatility slices are not guaranteed to be arbitrage free.

²³We have decided to include the beginning of the financial crisis so that the in-sample period actually includes several dates with extreme events.

values). The last column of Table 4 presents the number of options within each moneyness and maturity range in both periods. In particular, our dataset contains, in the in-sample period, 608 close-to-maturity OTM call options on the S&P 500 with a moneyness larger than 1.05, and 2,243 OTM put options with a moneyness smaller than 0.95. These options have maturities shorter than two months. Analogously, in the out-of-sample period, the dataset contains 737 close-to-maturity OTM calls and 2,032 OTM put options. As highlighted in Bollerslev and Todorov (2011), these options provide valuable information on jumps as they have little value unless a large movement in the S&P 500 is possible. Similarly, OTM calls on the VIX with short maturities 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 contains 1,006 such options on the VIX with a moneyness larger than 1.1 in the in-sample period and 1,269 in the out-of-sample period.

G Variance risk premium

The expression of the VRP involves the following terms:

$$\mathbb{E}^{\mathbb{Q}} \left(\int_t^T v_s ds \right) = A(T-t)v_t + B(T-t)m_t + D(T-t)u_t + G(T-t); \quad (\text{G.1})$$

$$\mathbb{E}^{\mathbb{Q}} \left(\int_t^T m_s ds \right) = \hat{A}(T-t)m_t + \hat{B}(T-t); \quad (\text{G.2})$$

$$\mathbb{E}^{\mathbb{Q}} \left(\int_t^T u_s ds \right) = \bar{A}(T-t)u_t + \bar{B}(T-t); \quad (\text{G.3})$$

$$(\text{G.4})$$

where $A, B, D, G, \hat{A}, \hat{B}, \bar{A}$ and \bar{B} are given by (A.6)-(A.9). Expectations under \mathbb{P} are calculated following the same procedure.

To calculate the discontinuous component of the VRP, we use that:

$$\mathbb{E}_t^{\mathbb{Q}} \left(\sum_{i=N_t^{(+/-)}}^{N_T^{(+/-)}} \left(Z_i^{(+/-)} \right)^2 \right) = \mathbb{E}_t^{\mathbb{Q}} \left(\sum_{i=N_t^{(+/-)}}^{N_T^{(+/-)}} 2 \left(\mu_Y^{(+/-)} \right)^2 \right) \quad (\text{G.5})$$

$$= 2 \left(\mu_Y^{(+/-)} \right)^2 \left(\lambda_0^{(+/-)} (T - t) + \lambda_1^{(+/-)} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T v_s ds \right] + \lambda_2^{(+/-)} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T m_s ds \right] + \lambda_3^{(+/-)} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T u_s ds \right] \right). \quad (\text{G.6})$$

Table 1. Descriptive statistics for daily S&P 500 futures log-returns and daily VIX levels for the periods from March 2006 to February 2009, and March 2009 to October 2010. We report the mean (Mean), standard deviation (Std), skewness (Skew), and kurtosis (Kurt).

	March 2006–February 2009				March 2009–October 2010			
	Mean	Std	Skew	Kurt	Mean	Std	Skew	Kurt
S&P 500	-0.0007	0.0159	0.0722	14.0772	0.0007	0.0158	-0.3283	6.8683
VIX	0.2044	0.1208	2.6560	10.9620	0.2907	0.1025	1.1929	3.8631

	\mathcal{D}_1			\mathcal{D}_2		
	SV2	SVJ	SVJ2	SVJ3	SVJ2	SVJ3
LL		8713	9200	9412	7705	8850
AIC		-17384	-18346	-18766	-15348	-17634
BIC		-17290	-18225	-18636	-15209	-17486

Table 2. Log-likelihood (LL) and values of the AIC and BIC tests for sub-models of the SVJ3 models. The specifications considered are: no jumps (SV2), m and u are constant (SVJ), u is constant (SVJ2). The estimation period is from March 2006 to February 2009.

Table 3. Parameters of the SVJ3 model, estimated using \mathcal{D}_1 and \mathcal{D}_2 as estimation datasets. The standard errors are in italics below each parameter, except for θ_m^P , which is calculated as $\theta_m^Q = \theta_m^Q \kappa_m^Q / \kappa_m^P$. \mathcal{D}_1 contains S&P 500 and VIX index levels and S&P 500 options. \mathcal{D}_2 contains in addition VIX options. The log-likelihood (AIC), and AIC and BIC test values are reported below the parameters. The estimation period is from March 2006 to February 2009.

	\mathbb{P} and \mathbb{Q} parameters		\mathbb{P} parameters		\mathbb{Q} parameters	
	\mathcal{D}_1	\mathcal{D}_2	\mathcal{D}_1	\mathcal{D}_2	\mathcal{D}_1	\mathcal{D}_2
$\lambda_1^{Y(-)}$	1.05	1.70	$\kappa_v^{\mathbb{P}}$	7.40	$\kappa_v^{\mathbb{Q}}$	5.30
	<i>0.07</i>	<i>0.06</i>		<i>0.15</i>		<i>0.07</i>
$\lambda_2^{Y(-)}$	0.20	0.25	$\kappa_m^{\mathbb{P}}$	0.13	$\kappa_m^{\mathbb{Q}}$	0.24
	<i>0.05</i>	<i>0.03</i>		<i>0.00</i>		<i>0.02</i>
$\lambda_3^{Y(-)}$	0.02	0.03	$\nu_v^{\mathbb{P}}$	0.02	$\theta_m^{\mathbb{Q}}$	0.02
	<i>0.01</i>	<i>0.01</i>		<i>0.00</i>		<i>0.01</i>
$\lambda_1^{Y(+)}$	1.30	0.20	$\mu_Y^{\mathbb{P}}$	-0.04	$\nu_v^{\mathbb{Q}(+)}$	0.05
	<i>0.02</i>	<i>0.03</i>		<i>0.00</i>		<i>0.00</i>
$\lambda_2^{Y(+)}$	1.00	0	$\mu_Y^{\mathbb{P}}$	0.02	$\mu_Y^{\mathbb{Q}(-)}$	-0.11
	<i>0.12</i>			<i>0.00</i>		<i>0.00</i>
σ_m	0.10	0.14	η_Y	0.60	$\mu_Y^{\mathbb{Q}(+)}$	0.00
	<i>0.00</i>	<i>0.01</i>		<i>0.04</i>		<i>0.01</i>
σ_v	0.57	0.69				
	<i>0.02</i>	<i>0.02</i>				
σ_u	0.25	0.30				
	0.00	<i>0.03</i>				
ρ_{Yv}	-0.80	-0.74				
	<i>0.01</i>	<i>0.02</i>				
κ_u	3.30	4.00				
	0.12	<i>0.12</i>				
α_0	0.30	0.30				
	<i>0.01</i>	<i>0.03</i>				
α_1	0.65	0.65				
	<i>0.01</i>	<i>0.03</i>				
$\nu_v^{(-)}$	-0.04	-0.06				
	<i>0.00</i>	<i>0.01</i>				
LL	9412	8850				
AIC	-18766	-17634				
BIC	-18636	-17486				

Table 4. Root Mean Square Relative Errors (RMSRE) on the implied volatilities of S&P 500 options using the SVJ3 model with estimation datasets \mathcal{D}_1 and \mathcal{D}_2 . Option pricing errors are reported conditional on moneyness (Mon.) and time to maturity (TTM) in months (M). The last column reports the number of options (#) in each category. Moneyness is defined as the ratio between the strike and the corresponding maturity future's price.

	\mathcal{D}_1	\mathcal{D}_2	#	\mathcal{D}_1	\mathcal{D}_2	#
	In-sample			In-sample until September 2008		
Overall	0.096	0.099	15952	0.075	0.079	13077
$Mon. < 0.7$	0.139	0.097	1776	0.081	0.046	1147
$0.7 \leq Mon. \leq 0.95$	0.084	0.059	6969	0.078	0.055	6098
$0.95 \leq Mon. \leq 1.05$	0.073	0.108	3848	0.067	0.108	3497
$1.05 \leq Mon. \leq 1.2$	0.022	0.021	2557	0.018	0.019	2048
$Mon. > 1.2$	0.149	0.208	802	0.038	0.097	287
$TTM \leq 2 M$	0.111	0.121	4997	0.087	0.100	4087
$2 M \leq TTM < 6 M$	0.086	0.089	6069	0.064	0.069	5004
$TTM > 6 M$	0.090	0.084	4886	0.074	0.067	3986

(a)

	Out-of-sample period	
Overall	0.130	0.108
$Mon. < 0.7$	0.946	0.946
$0.7 \leq Mon. \leq 0.95$	0.126	0.087
$0.95 \leq Mon. \leq 1.05$	0.071	0.068
$1.05 \leq Mon. \leq 1.2$	0.082	0.103
$Mon. > 1.2$	0.075	0.152
$TTM \leq 2 M$	0.132	0.110
$2 months \leq TTM < 6 M$	0.125	0.101
$TTM > 6M$	0.136	0.116

(b)

Table 5. Root Mean Square Relative Errors (RMSRE) on the implied volatilities of VIX options using the SVJ3 model with estimation datasets \mathcal{D}_1 and \mathcal{D}_2 . Option pricing errors are reported conditional on moneyness (Mon.) and time to maturity (TTM) in months (M). The last column reports the number of options (#) in each category. Moneyness is defined as the ratio between the strike and the corresponding maturity future's price.

	(a)			(b)		
	\mathcal{D}_1	\mathcal{D}_2	#	\mathcal{D}_1	\mathcal{D}_2	#
	In-sample			In-sample until September 2008		
Overall	0.613	0.423	4768	0.528	0.354	3562
$Mon. > 1.3$	0.361	0.275	1312	0.134	0.073	4680
$1.1 \leq Mon. \leq 1.3$	0.366	0.220	1055	0.162	0.078	3120
$0.9 \leq Mon. \leq 1.1$	0.533	0.296	1205	0.464	0.193	965
$Mon. < 0.9$	0.976	0.729	1196	0.878	0.675	802
$TTM \leq 2 M$	0.636	0.541	2283	0.531	0.378	1662
$TTM > 2 M$	0.564	0.301	2485	0.505	0.255	1900
	Out-of-sample period					
Overall			0.653	0.490		4498
$Mon. > 1.3$			0.476	0.381		1827
$1.1 \leq Mon. \leq 1.3$			0.519	0.334		824
$0.9 \leq Mon. \leq 1.1$			0.676	0.413		979
$Mon. < 0.9$			0.980	0.799		868
$TTM \leq 2 M$			0.624	0.484		868
$TTM > 2 M$			0.673	0.494		1902

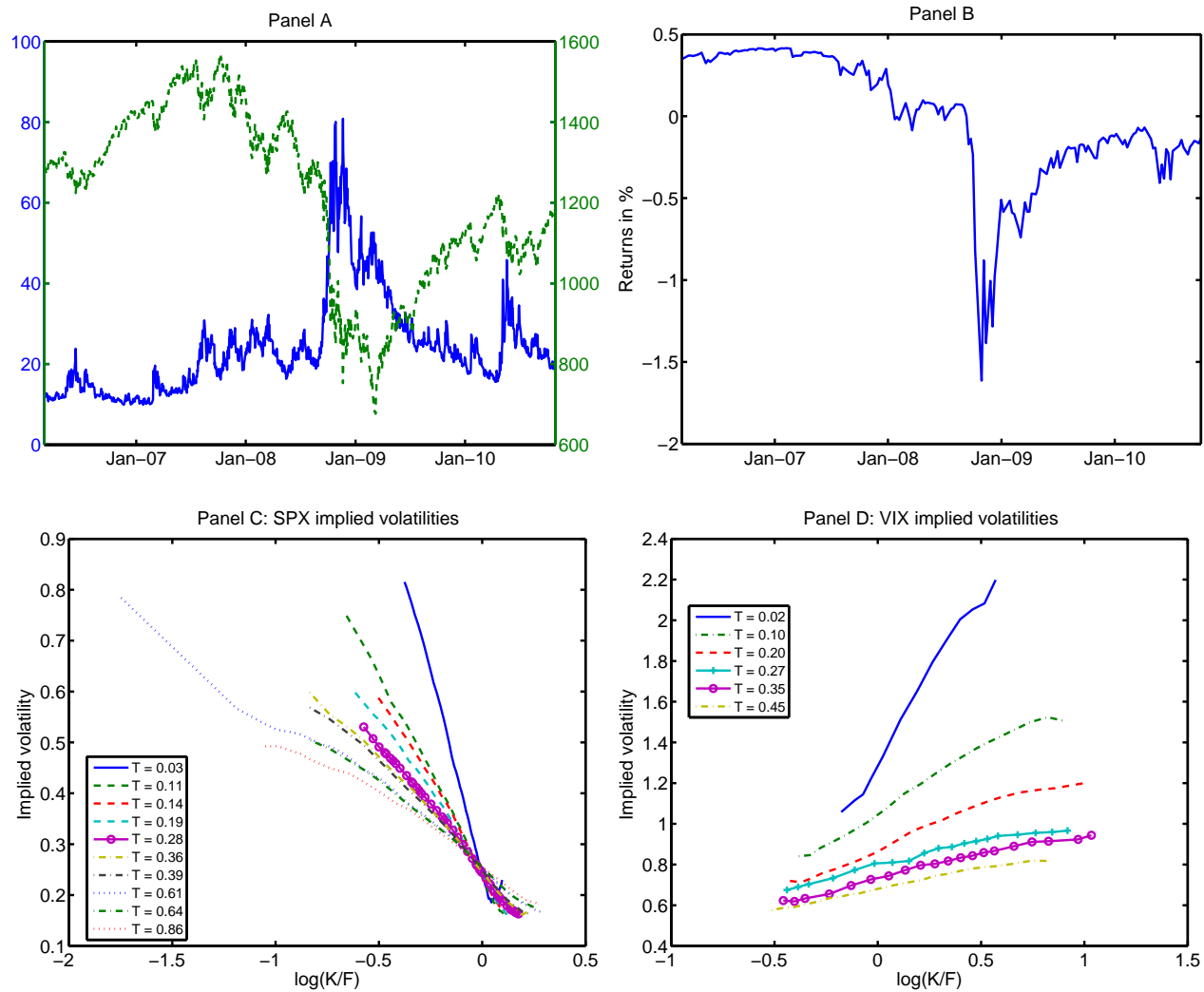


Figure 1. Panel A: Times series of S&P 500 (dashed curve) and VIX (solid curve) indexes from March 1st, 2006 to October 29th, 2010. Panel B: One-month S&P 500 expected returns implied by S&P 500 options with one month maturity. We use the method described in Bakshi, Kapadia, and Madan (2003). Returns are expressed in percentages. Panels C and D: Implied volatilities of S&P 500 options and VIX options on May 10 2010, as a function of log-moneyness $\log(K/F)$. The maturities T are quoted in years.

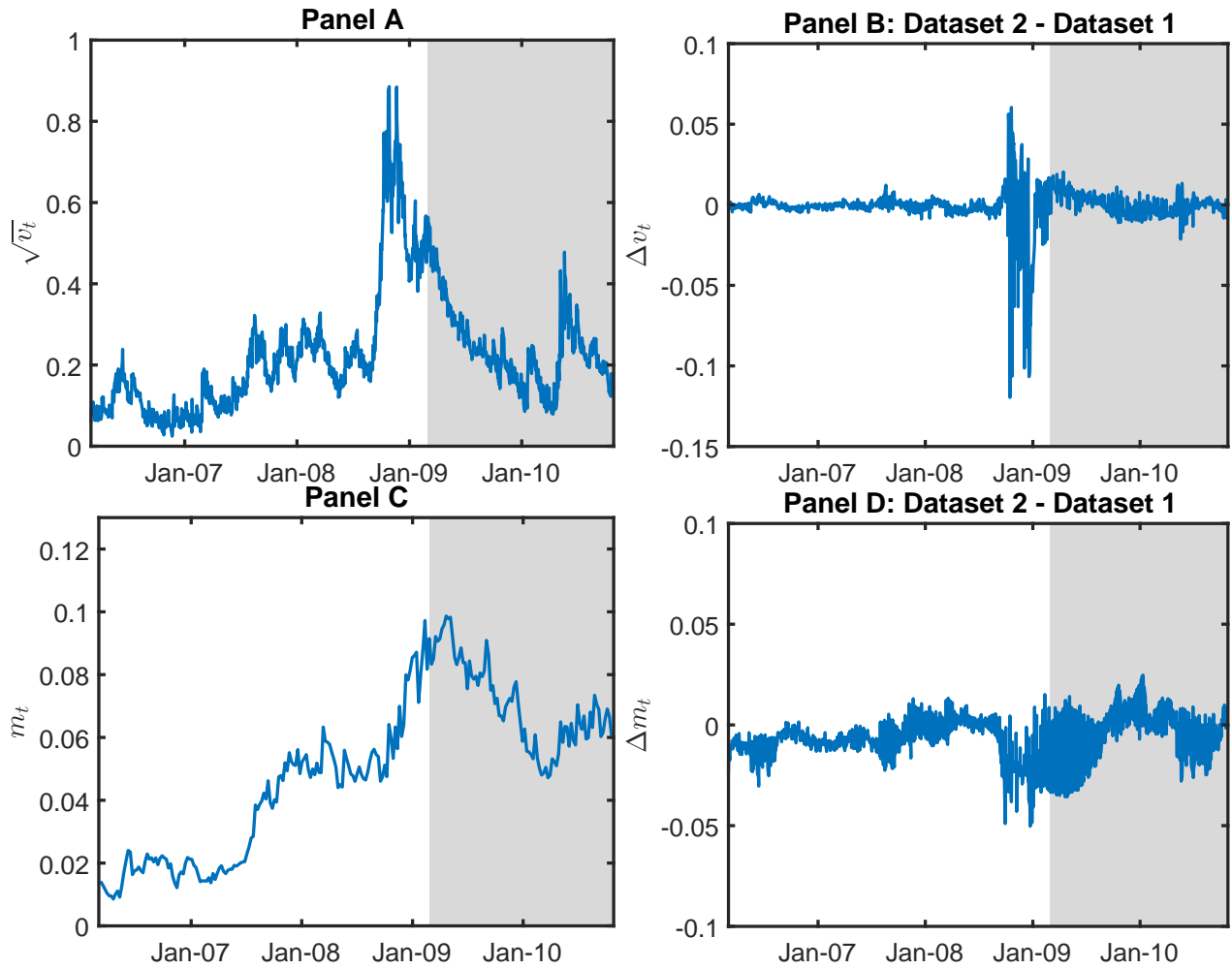


Figure 2. Panels A and C: Filtered trajectories of the latent volatility processes v and m for the SVJ3 model using \mathcal{D}_2 from March 2006 to March 1st, 2009 (748 days). Panels B and D: difference between the filtered v and m processes of the SVJ3 model for datasets \mathcal{D}_2 and \mathcal{D}_1 . The shaded part of the graph represents the out-of-sample period, from March 1, 2009 until the end of October 2010.

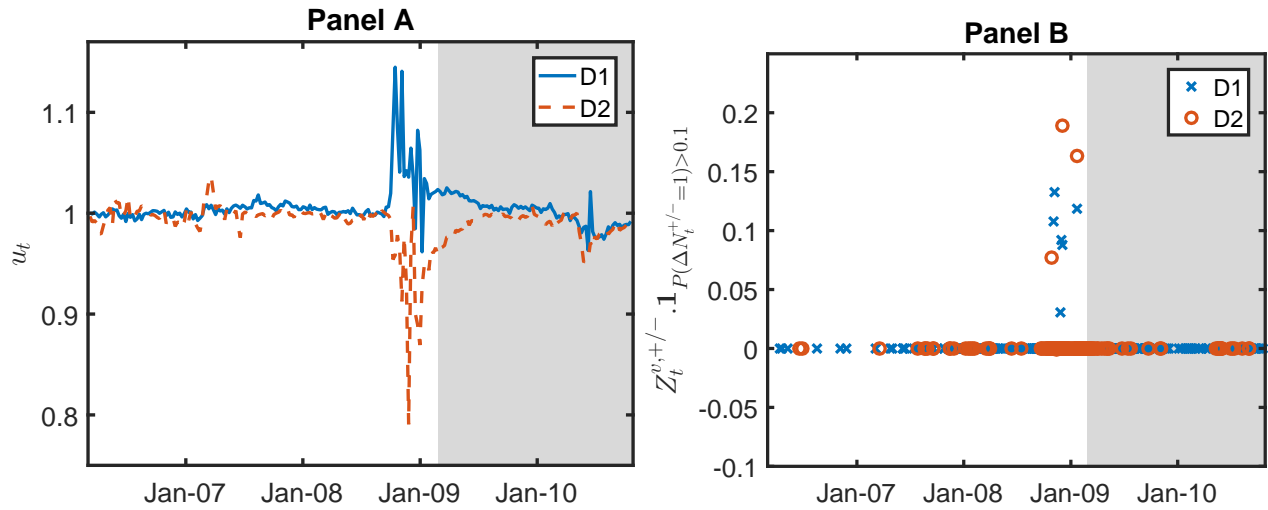


Figure 3. Panel A: Filtered trajectories of the latent jump intensity process u for the SVJ3 model, with and without VIX options in the estimation dataset, from March 2006 to March 1st, 2009 (748 days). Panel B: Filtered size of jumps in the variance process v , when the probability of jumps exceeds 10%. The shaded part of the graph represents the out-of-sample period, from March 1, 2009 until the end of October 2010.

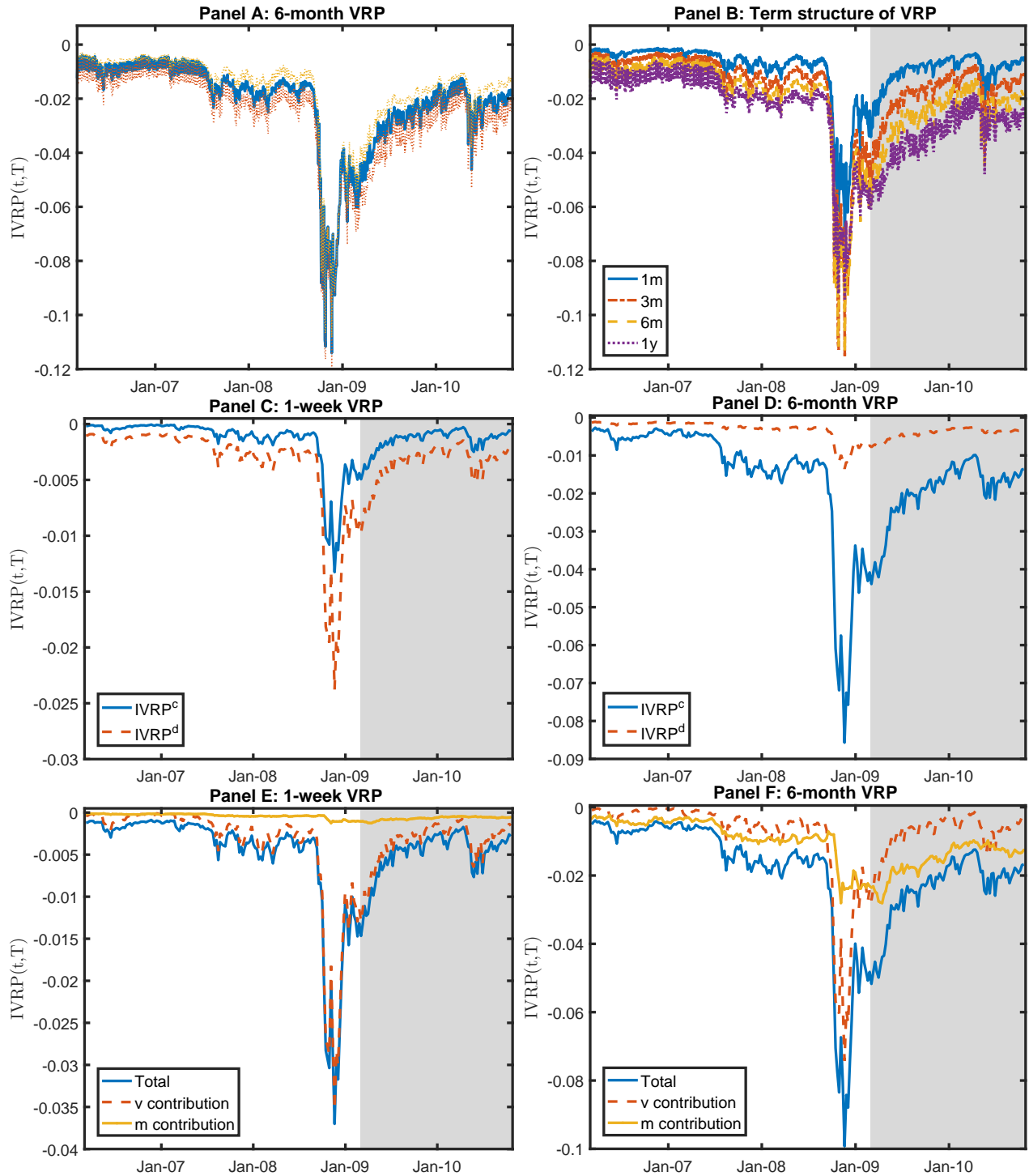


Figure 4. Integrated variance risk premia (VRP) when estimating the SVJ3 model using the full dataset of S&P 500 options, VIX options and their underlying levels. Panel A plots the 6-month VRP and its 90% confidence interval, conditional on parameter estimates. Panel B plots the VRP for different maturities. Panels C and D decompose the 1-week and 6-month VRP into their continuous and discontinuous components. Panels E and F plot the contribution of the latent factors v and m to the 1-week and 6-month VRP. The shaded parts of the graphs represent the out-of-sample period from March 1, 2009 until the end of October 2010.

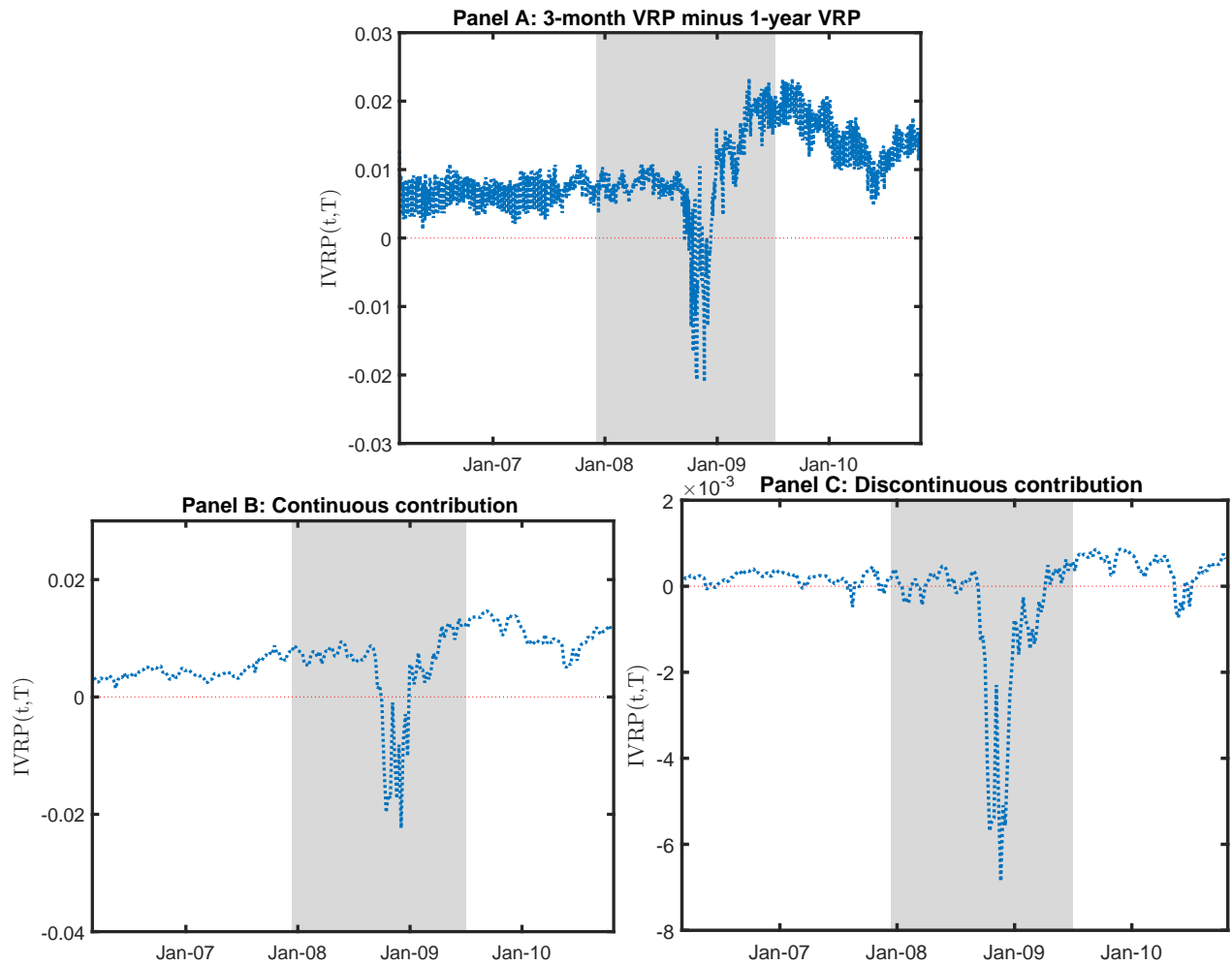


Figure 5. Difference between the 3-month and the 1-year integrated variance risk premia (VRP) when estimating the SVJ3 model using the full dataset of S&P 500 options, VIX options and their underlying levels. Panel B plots the contribution of the continuous fluctuations in the VRP and panel C the contribution of the jumps. The shaded part of the graphs represents the NBER recession from December 2007 until June 2009.

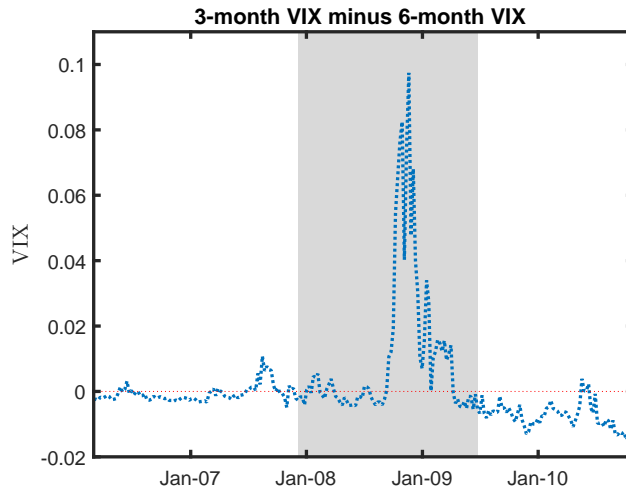


Figure 6. Difference between the 3-month and the 6-month VIX constructed from S&P 500 options. The shaded part of the graphs represents the NBER recession from December 2007 until June 2009.

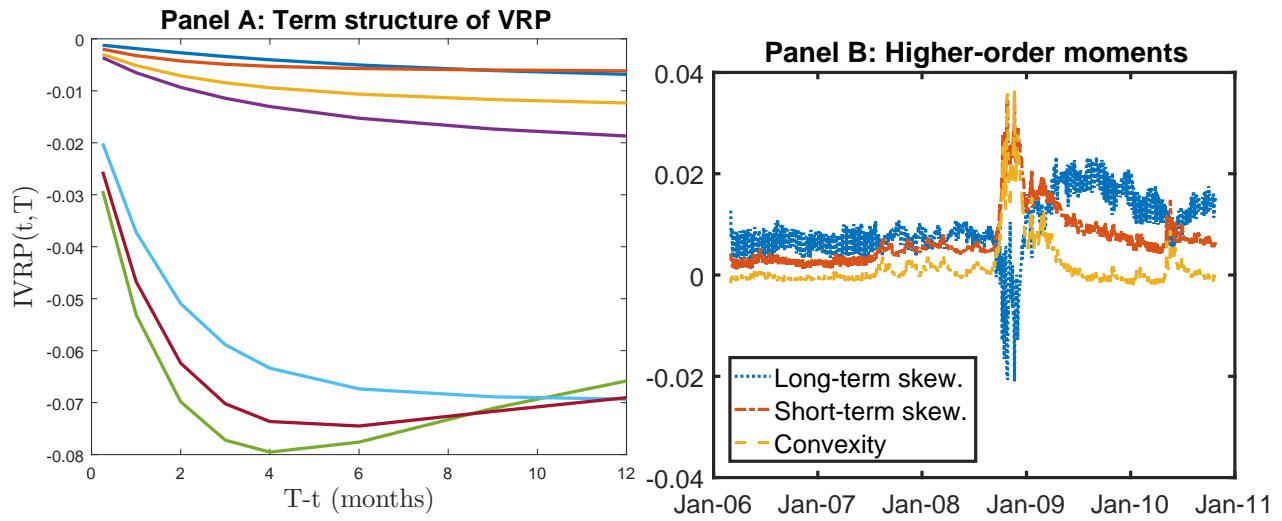


Figure 7. Panel A: Term structure of variance risk premium for investment horizons from 1 month to 1 year. The VRP is represented on the following dates, going from the upper curve to the lower one: 3 March 2006, 17 May 2006, 14 June 2006, 18 June 2008, 5 November 2008, 10 December 2008, 22 October 2008. Panel B: Long- and short-term skewness of the term structure of VRP over time.

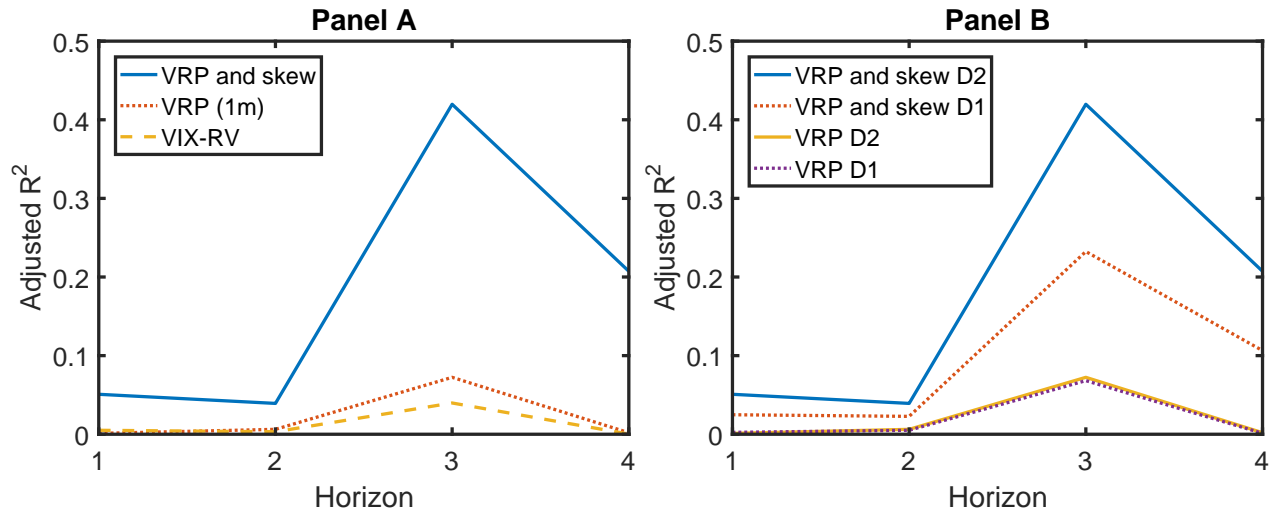


Figure 8. Adjusted R^2 when regressing future returns over the horizon given on the x-axis (from a 1-month to a 1-year horizon) over three sets of regressors: {the VIX minus the 1-month realized variance}, {the 1-month VRP} and {the 1-month VRP and the short-term skewness of the VRP term structure}.

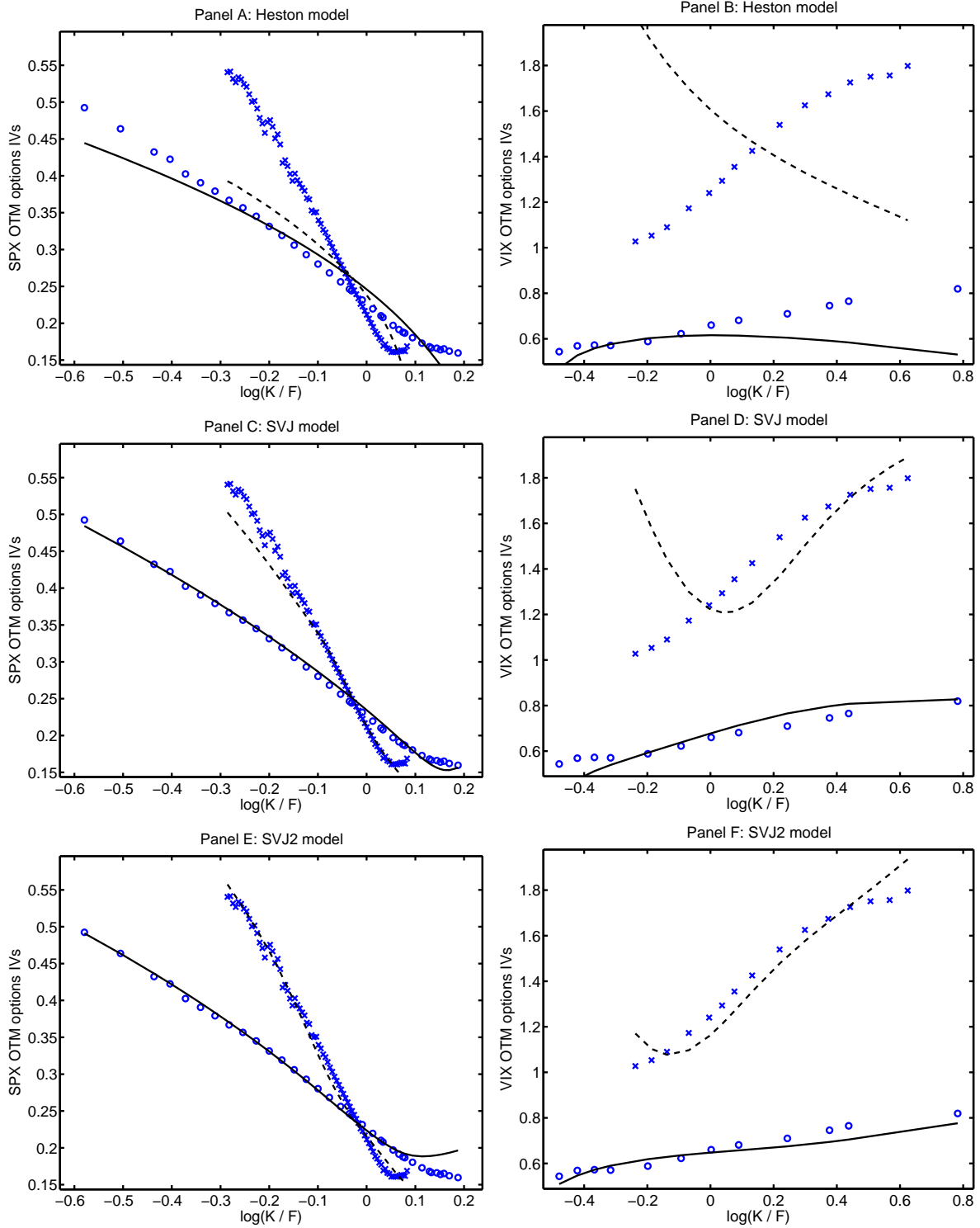


Figure 9. Market and model IVs on May 5, 2010, obtained by a joint calibration on the S&P 500 and VIX option market on that day. Circles represent the market IV for $T = 0.05$ (S&P 500) and $T = 0.04$ (VIX). Crosses represent the market IV for $T = 0.3$ (S&P 500) and $T = 0.36$ (VIX). The dashed line corresponds to the model fit for $T = 0.05$ (S&P 500) and $T = 0.04$ (VIX) while the solid line corresponds to the model fit for $T = 0.3$ (S&P 500) and $T = 0.36$ (VIX). Panels A (S&P 500) and B (VIX) plot the model IVs based on the Heston model. Panels C and D display the corresponding results for the SVJ model, while Panels E and F do so for the SVJ2 model.