Journal of Financial Economics 131 (2019) 593–618

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

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

Chris Bardgett[a], Elise Gourier[b], Markus Leippold[a,*]

[a] University of Zurich, Plattenstrasse 14, Zürich 8032, Switzerland
[b] ESSEC Business School and Centre for Economic Policy Research (CEPR), 3 Avenue Bernard Hirsch, 95021 Cergy-Pontoise, France

ARTICLE INFO

Article history:
Received 15 July 2015
Revised 22 January 2018
Accepted 6 February 2018
Available online 17 September 2018

JEL classification:
G12
G13
C58

Keywords:
S&P 500 and VIX joint modeling
Volatility dynamics
Particle filter
Variance risk premium

ABSTRACT

We estimate a flexible affine model using an unbalanced panel containing S&P 500 and VIX index returns and option prices and analyze the contribution of VIX options to the model's in- and out-of-sample performance. We find that they contain valuable information on the risk-neutral conditional distributions of volatility at different time horizons, which is not spanned by the S&P 500 market. This information allows enhanced estimation of the variance risk premium. We gain new insights on the term structure of the variance risk premium, present a trading strategy exploiting these insights, and show how to improve S&P 500 return forecasts.

1. Introduction

Introduced by the Chicago Board of Exchange (CBOE) in 1993, the VIX index nonparametrically 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 data sets 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

☆ We thank G. William Schwert (the editor), an anonymous referee as well as Yacine Aït-Sahalia, Torben Andersen, Tim Bollerslev, Peter Christoffersen, Rama Cont, Jérôme Detemple, Garland Durham, Damir Filipović, Andras Fulop, René Garcia, Michael Johannes, Loriano Mancini, Ian Martin, 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, Rotman School of Management, University of Montreal, UCL, LSE and CREST 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 2015 French Econometrics Conference at TSE, CFE 2016, the 2016 Paris Finance Meeting, and the 2017 Empirical Finance Workshop at ESSEC. 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. Finally, we are especially grateful for the detailed feedback of an anonymous referee. Any remaining errors are ours.

* Corresponding author.
E-mail addresses: chris.bardgett@bf.uzh.ch (C. Bardgett), elise.gourier@essec.edu (E. Gourier), markus.leippold@bf.uzh.ch (M. Leippold).

https://doi.org/10.1016/j.jfineco.2018.09.008

information on the dynamic properties of S&P 500 returns, which is not spanned by S&P 500 options and can be used for economic applications such as portfolio allocation or return forecasts. 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 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 et al. (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 data set 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 that 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 data set. This observation holds both in and out of sample, and is therefore not the result of overfitting. 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 allows a better representation of the term structure of variance. We synthesize VIX-type of indexes from S&P 500 options for maturities from two to six months. Our model, when estimated to a data set that does not include VIX options, yields root-mean-squared errors (RMSEs) that are 15–20% larger than the ones obtained when VIX options are included in the estimation data set. 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 VRP, which represents the compensation investors expect to receive for bearing the risk coming from stochastic fluctuations in the variance of returns over a given horizon. The VRP corresponds to 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. Hence, large movements in the variance process have 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 et al. (2015) and Dew-Becker et al. (2017). The latter authors 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. In line with them, 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. 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 et al. (2017) 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 toward long-term volatility changes over time and depends on market conditions.

While the usual trading strategy to reap the VRP is to buy long-term variance contracts and sell short-term contracts,[1] we show that the change in the term structure of VRP can be interpreted as a trading signal to improve the gains of this strategy. Indeed, the usual long-short strategy suffers sizable punctual losses in periods of high volatility, i.e., precisely when the term structure of VRP switches sign. We propose to switch the sign of the position in forward variance as soon as the VRP term structure switches sign. Our proposed strategy achieves a Sharpe ratio of 0.77

[1] To implement such a strategy, one can use variance swaps, see e.g., Egloff et al. (2010) and Filipović et al. (2016).

over the period from 2006 to 2016, against 0.01 for the usual variance swap strategy.

Finally, we examine the predictive power of our model-implied VRP on S&P 500 returns. We find that the term structure of the VRP has predictive power on S&P 500 returns as well, in particular for horizons larger than five months. Indeed, adding a measure of skewness of the VRP term structure (or, equivalently, of convexity) to the model-implied VRP level increases the R^2 from 0.15 (0.12) to 0.16 (0.15) for a horizon of five (seven) months. This result is intuitive. Indeed, we show that the dynamics of the VRP are well described by two latent factors. Having the VRP level and a measure of its term structure allows spanning these two factors, which are shown to contain relevant information on future returns.

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 and 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. However, we perform all tests in the paper using a benchmark two-factor affine model; all our results hold using both specifications.

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 data set 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 data set 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 that 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 et al. (2011), Cont and Kokholm (2013), Papanicolaou and Sircar (2014), and Bayer et al. (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 et al. (2000) that includes as subcases the usual extensions of the Heston model encountered in the literature, for example Bates (2000), Eraker (2004), and Sepp (2008).[2] In related work, Song and Xiu (2016) 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. In particular, they find, interestingly, a pricing kernel of the VIX that is U-shaped, similarly to the kernel of market returns. Because our data set in liquidly traded VIX options mainly contains calls and therefore information on the right tail of the variance distribution, we remain agnostic about the price of large downward volatility changes. In contrast to their paper where they focus on prices of risks over a 42-day horizon, we analyze the added information content of VIX options on the entire term structure of variance and the subsequent economic implications in terms of portfolio allocation and return forecasts.

We also build on a literature that 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 et al. (2015), solely based on S&P 500 options. Gruber et al. (2015) differentiate between periods of low and high volatility, and Dew-Becker et al. (2017) 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. Our results are complementary to theirs.

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 (2000), Pan (2002), Eraker (2004), Broadie et al. (2007), Christoffersen et al. (2010), Johannes et al. (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 contrast, in our particle filter estimation we fully exploit the richness of our data set. Furthermore, we note that most 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

[2] Some studies are going in the direction of nonaffine models (Jones, 2003; Aït-Sahalia and Kimmel, 2007; Christoffersen et al., 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.

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.[3]

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 expressions of the VRP and of the VIX squared as well as the pricing formula for VIX and S&P 500 options. In Section 3, we describe our data set. 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 and the properties of its term structure. In Section 7, we discuss two economic implications of our model and estimation methodology. The first one shows how the model-implied term structure of VRP can be used as a trading signal, and the second one examines the predictive power of the VRP on future S&P 500 returns. Section 8 concludes.

2. Theoretical framework

We first present our modeling framework. Our model is novel and able to capture 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 et al. (2015) and Amengual and Xiu (2015), which makes it possible to better capture the stochastic skewness of returns and of their variance. Despite its flexibility, it is parsimonious and tractable, as it belongs to the affine model class.

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

The process $v = (v_t)_{t\geq 0}$ is the diffusive component of the variance of the S&P 500 returns. The second variance factor $m = (m_t)_{t\geq 0}$ represents a stochastic level around which v reverts.[4] We refer to it as central tendency. The increments of the two processes v and m are instantaneously uncorrelated, and v and m 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

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

where $Z_i^{Y(+)}$, $Z_i^{Y(-)}$, $Z_i^{v(+)}$, and $Z_i^{v(-)}$ represent the random jump sizes. As suggested by the price paths of the S&P 500 and VIX index, large negative (positive) movements in equity returns and large positive (negative) movements in the variance are likely to occur at the same time. We therefore choose, in line with 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 volatility, accounting for negative jumps has been less investigated. 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 levels of the diffusive latent processes v, m, and u.[5] The intensity of positive jumps in returns is denoted by $\lambda^{(+)}$ and the intensity of negative jumps by $\lambda^{(-)}$:

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

where $\mathbf{X}$ denotes the (column) vector of diffusive 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 is an additional driver of the intensity of negative jumps in returns (together with positive jumps in variance), as in Andersen et al. (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 us to better represent the stochastic skewness of the return process. Andersen et al. (2015) find that

[3] See, e.g., Lindström et al. (2008).

[4] It has already been shown that at least two factors are needed to provide an accurate description of the volatility dynamics (Bates, 2000; 2012; Andersen et al., 2002; Alizadeh et al., 2002; Chernov et al., 2003; Christoffersen et al., 2009; Egloff et al., 2010; Todorov, 2010; Kaeck and Alexander, 2012; Johnson, 2012; Mencía and Sentana, 2013; Huang and Shaliastovich, 2015; Branger et al., 2016).

[5] The specification of jumps is of importance. Todorov (2010), Todorov and Tauchen (2011), and Jacod and Todorov (2010) find striking evidence for cojumps in S&P 500 returns and in the VIX. See also Eraker (2004), Broadie et al. (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 et al. (2012) find that the state-dependent intensity of jumps is a desirable model feature.

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. For the jumps in return, we assume that positive and negative jumps are exponentially distributed with respective means $\mu_Y^{(+)}$ and $\mu_Y^{(-)}$.

The positive and negative jump sizes in return volatility are assumed to be exponentially distributed with mean $\nu_\nu^{(+)}$ and $\nu_\nu^{(-)}$, respectively. Let us define $\mathbf{Z}_i = (Z_i^{Y(+)}, Z_i^{Y(-)}, Z_i^{\nu(+)}, Z_i^{Y(-)})^\top$, $i \in \mathbb{N}^*$. The jump sizes are characterized by their joint Laplace transform

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

The drift of the returns process can be written accordingly as

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

Since with the above specification, the model is driven by three latent processes ν, m, and u, we refer to its general form as the SVJ3 model. In this SVJ3 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 ν. While this is certainly a drawback, reaching zero is already possible with a standard Heston model with positive jumps only, when the Feller condition is not satisfied. Song and Xiu (2016), among others, find that the Feller condition is violated by the data. To tackle this issue, we verify in the empirical part of the paper that the estimated trajectory of the process ν_t never touches or crosses the zero boundary.

The above 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 $\mathrm{VIX}_t^2 = \frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}[\int_t^{t+\tau} \frac{dF_u}{F_{u^-}} - d(\ln F_u)]$, where τ is 30 days in annual terms.

Proposition 2.1. Under the model specification given in Eqs. (1)–(7), the VIX squared at time t can be written as an affine deterministic function of ν_t, m_t, and u_t:

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

$$= \alpha_{\mathrm{VIX}^2}\nu_t + \beta_{\mathrm{VIX}^2} m_t + \gamma_{\mathrm{VIX}^2} u_t + \delta_{\mathrm{VIX}^2}, \tag{10}$$

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. Benchmark model specification

To challenge our SVJ3 model, we specify a two-factor affine jump diffusion model as a benchmark. The dynamics of Y under $\mathbb{Q}$ are simplified to

$$dY_t = \mu(\nu_{t^-}, m_t)dt + \sqrt{\nu_{t^-}}dW_t^Y + dJ_t^Y,$$
$$d\nu_t = \kappa_\nu(m_t - \nu_{t^-})dt + \sigma_\nu\sqrt{\nu_{t^-}}dW_t^\nu + dJ_t^\nu, \tag{11}$$
$$dm_t = \kappa_m(\theta_m - m_t)dt + \sigma_m\sqrt{m_t}dW_t^m.$$

In this specification, we assume that jumps in returns are normally distributed $\mathcal{N}(\mu_Y, \sigma_Y)$, and volatility does not exhibit negative jumps. The intensity of jumps loads on ν and m but no longer loads on a separate u process. We refer to the specification in Eq. (11) as benchmark model for our SVJ3 model. This model corresponds to the benchmark model used in Filipović et al. (2016), which they estimate using variance swaps. It subsumes many of the popular stochastic volatility models as special cases, such as Bakshi et al. (1997), Bates (2000, 2006), Pan (2002), Eraker et al. (2003), and Broadie et al. (2007, 2009).

2.3. Risk premium specification

We specify the change of measure from the pricing to the historical measure so that the model dynamics have the same structure under $\mathbb{P}$. The premium for equity risk γ_t consists of 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:

$$\gamma_t = \eta_Y\nu_{t^-} + \lambda^{(+)}(\nu_{t^-}, m_t)\left(\theta_Z^{\mathbb{P}}(1,0,0,0) - \theta_Z(1,0,0,0)\right) + \lambda^{(-)}(\nu_{t^-}, m_t, u_t)\left(\theta_Z^{\mathbb{P}}(0,1,0,0) - \theta_Z(0,1,0,0)\right), \tag{12}$$

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 $\mathbb{P}$.[6] However, we allow the mean of the jump sizes in returns to be different under $\mathbb{Q}$ and $\mathbb{P}$.

Similarly, the instantaneous premium for volatility risk decomposes into a diffusive component and a jump component, for each of the volatility factors ν and m. The diffusive premium in ν is proportional to the current level of variance, with coefficient of proportionality given by $\eta_\nu = \kappa_\nu - \kappa_\nu^{\mathbb{P}}$. 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 premium, we allow the mean jump sizes $\nu_\nu^{(+)}$ and $\nu_\nu^{(-)}$ to be different under $\mathbb{P}$ and $\mathbb{Q}$.

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

2.4. Integrated variance risk premium

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

$$\mathrm{VRP}(t,T) = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(QV_{[t,T]}\right) - \mathbb{E}_t^{\mathbb{Q}}\left(QV_{[t,T]}\right)\right], \tag{13}$$

[6] Pan (2002) argues that introducing different intensities of jumps under the historical and pricing measure introduces a jump-timing risk premium that is very difficult to disentangle from the mean jump risk premium. Our assumption artificially incorporates the jump-timing risk premium into the mean jump size risk premium.

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:

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

The VRP represents the expected payoff when buying a variance swap at time t with maturity T. Alternatively, it reflects the amount investors are willing to pay for a hedge against future stochastic fluctuations in the variance. We can further decompose the VRP into a continuous and a discontinuous part:

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

with

$$\mathrm{VRP}^c(t,T) = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\int_t^T v_{s^-} ds\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T v_{s^-} ds\right)\right],$$

$$\mathrm{VRP}^d(t,T) = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\sum_{i=N_t^{(+)}}^{N_T^{(+)}} \left(Z_i^{(+)}\right)^2 + \sum_{i=N_t^{(-)}}^{N_T^{(-)}} \left(Z_i^{(-)}\right)^2\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\sum_{i=N_t^{(+)}}^{N_T^{(+)}} \left(Z_i^{(+)}\right)^2 + \sum_{i=N_t^{(-)}}^{N_T^{(-)}} \left(Z_i^{(-)}\right)^2\right)\right].$$

Each part can also be decomposed into linear contributions from m_t, v_t, and u_t, given that the expectations of the integrated latent factors are affine in their current values; see Appendix B.

2.5. 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 SVJ3 model defined by Eqs. (1)–(7), the Laplace transforms of VIX^2 *and the S&P 500 returns are exponential affine in the current values of the factor processes v, m and u:*

$$\Psi_{\mathrm{VIX}_T^2}(t,\tilde{x};\omega) := \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega \mathrm{VIX}_T^2}\middle|\mathbf{X}_t = \tilde{x}\right] = e^{\alpha(T-t)+\mathbf{B}(T-t)\tilde{x}},$$

$$\Psi_{Y_T}(t,y,\tilde{x};\omega) := \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega Y_T}\middle| y_t = y, \mathbf{X}_t = \tilde{x}\right] = e^{\alpha_Y(T-t)+\beta_Y(T-t)y+\mathbf{B}_\mathbf{Y}(T-t)\tilde{x}},$$

where α, α_Y, β_Y, $\mathbf{B} = (\beta,\gamma,\delta)^\top$, *and* $\mathbf{B}_\mathbf{Y} = (\gamma_Y,\delta_Y,\xi_Y)^\top$ *are functions defined on* $[0, T]$ *by the ordinary differential equations (ODEs) given in Appendix C. 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_{VIX} and maturity T on the VIX at time $t=0$, we need to calculate

$$C(\mathrm{VIX}_0, K_{\mathrm{VIX}}, T) = e^{-rT}\int_0^\infty (\sqrt{v} - K_{\mathrm{VIX}})^+ f_{\mathrm{VIX}_T^2}(v)dv, \quad (14)$$

where $f_{\mathrm{VIX}_T^2}$ 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 Eq. (14) 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 (2008) and Song and Xiu (2016), 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 $u_{\mathrm{VIX}}(\mathrm{VIX}^2) = (\sqrt{\mathrm{VIX}^2} - K_{\mathrm{VIX}})^+$. *Given an interval* $[a_{\mathrm{VIX}}, b_{\mathrm{VIX}}]$ *for the support of the* $\mathrm{VIX}_T^2|_{\mathbf{X}_0}$ *density, the price* $P_{\mathrm{VIX}}(t_0, \mathrm{VIX}_0)$ *at time* $t = t_0 \geq 0$ *of the contingent claim is approximated by*

$$P_{\mathrm{VIX}}(t_0, \mathrm{VIX}_0) \approx e^{-r(T-t_0)} {\sum_{n=0}^{N-1}}' A_n^{\mathrm{VIX}^2} U_n^{\mathrm{VIX}^2}, \quad (15)$$

where the prime superscript in the sum $\sum'$ *means that the first term* $A_0^{\mathrm{VIX}^2}U_0^{\mathrm{VIX}^2}$ *is divided by two, and N is a truncation threshold. The terms in the sum are defined by*

$$A_n^{\mathrm{VIX}^2} = \frac{2}{b_{\mathrm{VIX}} - a_{\mathrm{VIX}}} Re\left\{\Psi_{\mathrm{VIX}_T^2}\left(t_0, \mathbf{X}_0; \frac{in\pi}{b_{\mathrm{VIX}} - a_{\mathrm{VIX}}}\right) \times \exp\left(-ia_{\mathrm{VIX}}\frac{n\pi}{b_{\mathrm{VIX}} - a_{\mathrm{VIX}}}\right)\right\}, \quad (16)$$

$$U_n^{\mathrm{VIX}^2} = \int_{a_{\mathrm{VIX}}}^{b_{\mathrm{VIX}}} u_{\mathrm{VIX}}(v)\cos\left(n\pi\frac{v - a_{\mathrm{VIX}}}{b_{\mathrm{VIX}} - a_{\mathrm{VIX}}}\right)dv. \quad (17)$$

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

3. Data and preliminary analysis

In this section, we describe our data and point out some important characteristics of VIX options.

3.1. Data description

Options on the VIX were introduced in 2006. Our sample period is from March 1, 2006, to April 30, 2016. The option data consist of the weekly[7] 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.

Both the S&P 500 and VIX options data sets 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

[7] We follow Pan (2002) and Johannes et al. (2009), among others, and use weekly (Wednesday) options data. This eliminates beginning-of-week and end-of-week effects and reduces the computational burden of the estimation.

Table 1
Parameters of the SVJ3 model dynamics are given in Eqs. (1)–(3). The standard errors are in italics below each parameter, except for θ_m^P, which is calculated as $\theta_m^{\mathbb{P}} = \theta_m^{\mathbb{Q}}\kappa_m^{\mathbb{Q}}/\kappa_m^{\mathbb{P}}$. The estimation results in the columns "w/o" are based on S&P 500 and VIX index levels and S&P 500 options; those in the columns "w/" additionally use VIX options. The LL, 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	
	w/o	w/		w/o	w/		w/o	w/
$\lambda_1^{Y(-)}$	1.60	1.95	$\kappa_v^{\mathbb{P}}$	7.40	7.30	$\kappa_v^{\mathbb{Q}}$	5.60	5.00
	0.22	*0.21*		*0.22*	*0.12*		*0.21*	*0.18*
$\lambda_2^{Y(-)}$	1.60	1.90	$\kappa_m^{\mathbb{P}}$	0.13	0.24	$\kappa_m^{\mathbb{Q}}$	0.24	0.08
	0.77	*1.22*		*0.01*	*0.01*		*0.04*	*0.02*
$\lambda_3^{Y(-)}$	0.08	0.08	$\nu_v^{\mathbb{P}}$	0.02	0.03	$\theta_m^{\mathbb{Q}}$	0.02	0.04
	0.03	*0.05*		*0.01*	*0.02*		*0.01*	*0.01*
$\lambda_0^{Y(+)}$	0.03	0.09	$\mu_Y^{\mathbb{P}(-)}$	−0.04	−0.03	$\nu_v^{\mathbb{Q}(+)}$	0.05	0.07
	0.02	*0.09*		*0.01*	*0.02*		*0.04*	*0.01*
$\lambda_1^{Y(+)}$	1.30	1.02	$\mu_Y^{\mathbb{P}(+)}$	0.02	0.02	$\mu_Y^{\mathbb{Q}(-)}$	−0.11	−0.11
	0.16	*0.82*		*0.01*	*0.01*		*0.01*	*0.01*
σ_m	0.10	0.14	η_Y	0.60	0.58	$\mu_Y^{\mathbb{Q}(+)}$	0.02	0.04
	0.03	*0.01*		*0.14*	*0.05*		*0.02*	*0.01*
σ_v	0.65	0.69				LL	10,473	9949
	0.02	*0.02*				AIC	−20,898	−19,850
σ_u	0.35	0.30				BIC	−20,787	−19,739
	0.01	*0.03*						
ρ_{Yv}	−0.78	−0.74						
	0.01	*0.02*						
κ_u	2.50	4.00						
	0.27	*1.12*						
α_0	0.30	0.31						
	0.01	*0.03*						
α_1	0.65	0.65						
	0.02	*0.03*						
$\nu_v^{(-)}$	−0.04	−0.03						
	0.01	*0.01*						

traded on a given date. Then, we infer from highly liquid options the futures price using the at-the-money (ATM) put-call parity. By doing so, we avoid two issues: making predictions on future dividends and using futures closing prices that 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.[8]

These adjustments leave a total of 365,507 OTM S&P 500 and 44,539 call options on the VIX, with a daily average of 639 S&P 500 options and 78 VIX options. The number of S&P 500 (VIX) options in our data set on a given date increases with time, with around 170 (5) options at the beginning of the data set and around 2000 (160) 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 eight short maturities with a bit more than 20 options trading per maturity. For S&P 500 options, around 25 maturities are available per day with around 130 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. In 2010, the total VIX options volume per day is about half of the total volume of S&P 500 options traded, and at the end of our sample it is close to two-thirds. Options on both markets are hence liquidly traded.

3.2. Descriptive statistics

Table 1 presents the first four sample moments of the S&P 500 futures returns, VIX index returns, and square levels over two different periods of time. The first period starts in March 2006 and ends in February 2009, i.e., it spans the precrisis period as well as the beginning of the crisis. The second period begins in March 2009 and lasts until April 2016. For our estimation, these two periods serve as the in-sample and out-of-sample periods.

The S&P 500 returns exhibit a high kurtosis, especially during the in-sample period, suggesting the presence

[8] We 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 $[T, T+30d]$.

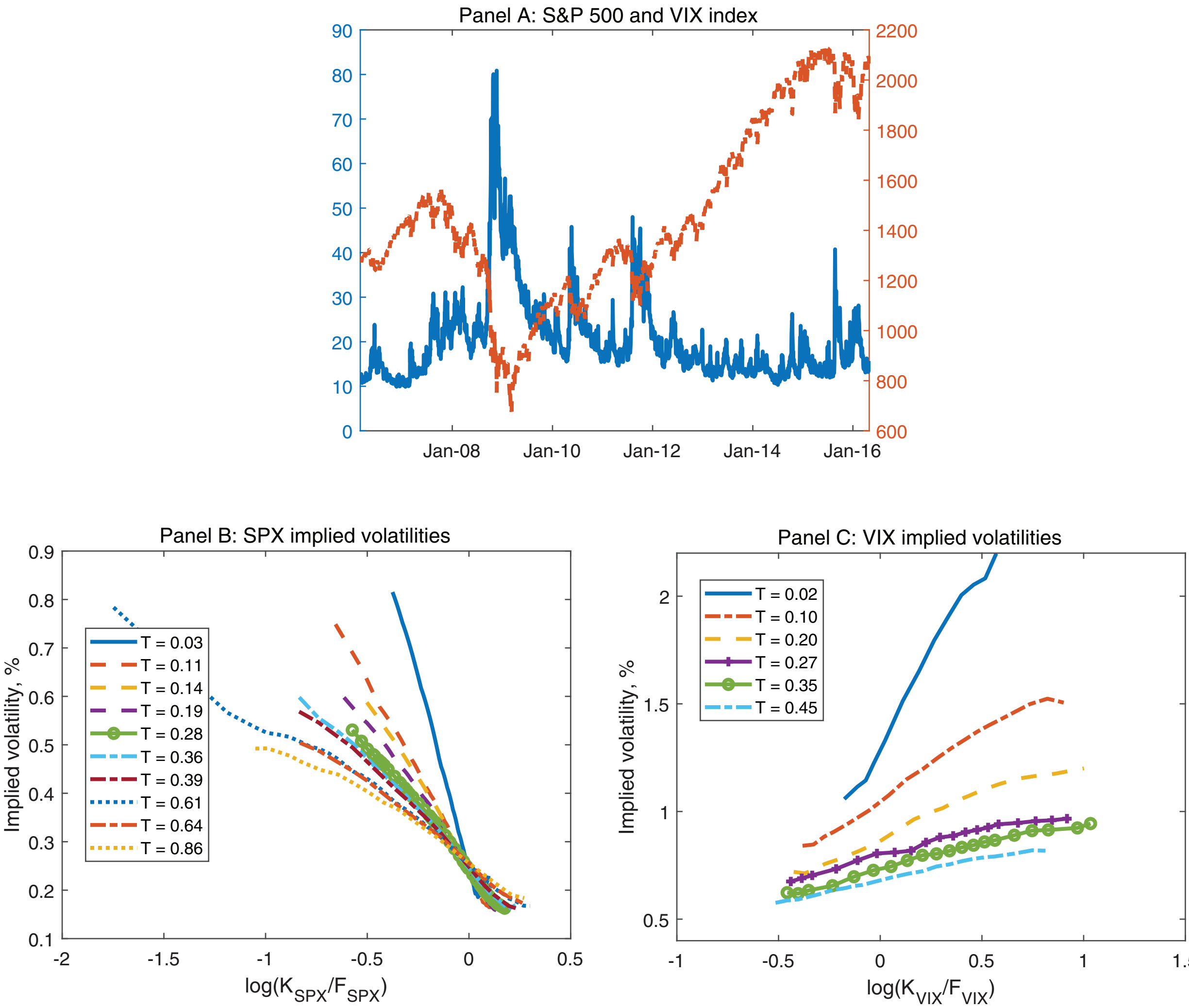

Fig. 1. This figure illustrates some characteristics of the S&P 500 and VIX markets. Panel A plots the times series of S&P 500 (dashed curve) and VIX (solid curve) indexes from March 1, 2006, to April 30, 2016. Panels B and C represent the implied volatilities of S&P 500 options and VIX options on May 10, 2010, as a function of log-moneyness. The maturities T are quoted in years.

of rare and large movements. In the in-sample period, their skewness is strongly negative due to the substantial losses made during the financial crisis. It remains slightly negative during the out-of-sample period. The VIX index exhibits a large positive skewness and kurtosis both in the in-sample and in the out-of-sample periods, although in the latter, both statistics decrease significantly.

Panel A of Fig. 1 displays the joint evolution of the S&P 500 and the VIX index from 2006 to 2016. The S&P 500 returns and the VIX daily increments are highly negatively correlated (with a correlation coefficient of -0.589 over this period), which explains the popularity of VIX contracts for hedging part of the equity risk of a portfolio.

These time series illustrate the variety of market situations covered by our time period. Both time series are rather stable until the summer of 2007. The S&P 500 grows almost steadily from 1200 to 1500, whereas the VIX level is around 10–15%. When the crisis of the quant-strategy hedge funds starts in the summer of 2007, the S&P 500 starts falling, whereas the VIX becomes more volatile and oscillates quickly around 20%. The bailout of Lehman Brothers in September 2008 accelerates the crash in the S&P 500 market, together with a large increase in the VIX index. The S&P 500 then starts a steady increase back to its original level and above, despite some punctual mini crashes. The VIX, in contrast, reacts much more drastically to new information. Following the bankruptcy of Lehman Brothers in September 2008 and the sudden increase to more than 80%, it goes down quickly but then up again when the House of Representatives rejects the Troubled Asset Relief Program at the end of September. It then reverts back to around 40% but increases 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 15% early 2010. The flash crash in May 2010 then leads to new heights, with a peak around 45%, and a reversion to around 10%. In 2011, the US debt downgrade, together with the Greek financial crisis, trigger a new peak in the VIX that is reversed following the agreement on a rescue plan. The last important peak in our time series is triggered by the Chinese yuan devaluation in May 2015.

Even though the S&P 500 and VIX markets are closely related, we emphasize that options on the VIX and S&P 500 substantially differ in their characteristics and in the information they contain on the underlying S&P 500 returns and variance. 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, they only provide us with a point estimate of the conditional variance of returns at each traded maturity. In turn, VIX options characterize the whole conditional density of future VIX levels. As such, they are likely to contain more information on the future density of S&P 500 return variance.

Panels B and C of Fig. 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 toward 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 B and C of Fig. 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 specific dates. 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. Our methodology and detailed results are reported in the online Internet Appendix, Section A.

To achieve a more robust estimation, 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.

4.1. Discretized model and specification of errors

The state space model is obtained by discretizing the continuous-time model under $\mathbb{P}$ on a uniform time grid in time. Measurements comprise the S&P 500 daily log-returns, the VIX levels, and the option prices on both indexes. The VIX and option prices are assumed to be observed with error. Indeed, 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 defining VIX^2. Regarding option prices, the error term represents several sources of noise, such as bid–ask spreads, timing, and processing errors.

To better identify the total variance of S&P 500 returns under the $\mathbb{P}$ measure, we add a measurement equation, which links the logarithm of the daily realized variance (RV_t) of S&P 500 returns[9] to the logarithm of the total spot variance under $\mathbb{P}$, as in Filipović et al. (2016). The associated measurement error ϵ_t is conditionally normally distributed with mean $\rho_\epsilon \epsilon_{t-1}$ and variance $c_0 + c_1 \mathrm{RV}_{t-1}$. The rationale behind this component of the measurement equation is the following. Andersen et al. (2001), among others, provide empirical evidence that $\log(\mathrm{RV}_t)$ 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 heteroskedasticity of the measurement error due to the volatility of realized variance. Details on the state space model are provided in the Internet Appendix, Section B.2.

9 The realized variance (RV) of the S&P 500 index is obtained from the website of the Oxford-Man Institute Realized Library.

Table 2

Parameters of the benchmark model dynamics given in Eq. (11). The standard errors are in italics below each parameter, except for $\theta_m^{\mathbb{P}}$, which is calculated as $\theta_m^{\mathbb{P}} = \theta_m^{\mathbb{Q}}\kappa_m^{\mathbb{Q}}/\kappa_m^{\mathbb{P}}$. The estimation results in the columns "w/o" are based on S&P 500 and VIX index levels and S&P 500 options; those in the columns "w/" additionally use VIX options. The estimation period is from March 2006 to February 2009.

	$\mathbb{P}$ - and $\mathbb{Q}$ -parameters			$\mathbb{P}$ -parameters			$\mathbb{Q}$ -parameters	
	w/o	w/		w/o	w/		w/o	w/
λ_0^Y	0.11	0.03	$\kappa_v^{\mathbb{P}}$	7.09	7.03	$\kappa_v^{\mathbb{Q}}$	5.36	5.34
	0.05	*0.04*		*0.20*	*0.18*		*0.15*	*0.19*
λ_1^Y	1.20	0.65	$\kappa_m^{\mathbb{P}}$	0.19	0.27	$\kappa_m^{\mathbb{Q}}$	0.17	0.21
	0.22	*0.31*		*0.04*	*0.05*		*0.09*	*0.09*
λ_2^Y	3.08	3.32	$\nu_v^{\mathbb{P}}$	0.02	0.01	$\theta_m^{\mathbb{Q}}$	0.04	0.05
	1.22	*1.19*		*0.00*	*0.02*		*0.01*	*0.01*
σ_m	0.20	0.21	$\mu_Y^{\mathbb{P}}$	0.00	−0.01	$\nu_v^{\mathbb{Q}}$	0.04	0.08
	0.05	*0.02*		*0.00*	*0.00*		*0.01*	*0.02*
σ_v	0.64	0.78	$\sigma_Y^{\mathbb{P}}$	0.03	0.04	$\mu_Y^{\mathbb{Q}}$	−0.08	−0.08
	0.03	*0.02*		*0.01*	*0.02*		*0.02*	*0.02*
ρ_{Yv}	−0.78	−0.78	η_Y	0.45	0.25	$\sigma_Y^{\mathbb{Q}}$	0.12	0.12
	0.02	*0.02*		*0.30*	*0.21*		*0.03*	*0.02*
α_0	0.30	0.35				LL	10,370	9538
	0.10	*0.14*				AIC	−20,704	−19,040
α_1	0.60	0.55				BIC	−20,621	−18,957
	0.04	*0.03*						

4.2. Particle filter

At every discrete point in time $t = t_n$, the measurement vector y_t collects observed market prices. By $y^t = (y_{t_0}, \ldots, y_{t_n})$, 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,

$$L_t = \left\{v_t, m_t, u_t, \Delta N_t^{(+)}, \Delta N_t^{(-)}, Z_t^{Y(+)}, Z_t^{Y(-)}, Z_t^{v(+)}, Z_t^{v(-)}\right\}, \tag{18}$$

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 et al. (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 the online Internet Appendix, Section B. In particular, we use the weighted likelihood method of Hu and Zidek (2002) to assign comparable weights 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.

5. Estimation results with and without VIX options

This section discusses the choice of the full model specification and analyzes how the model performs in representing the data, depending on whether we include VIX options in the estimation data set or not.

5.1. Model selection

Let us start by noting that options are crucial for identifying the parameters of our model. Even when estimating highly restricted subspecifications of our full specification to a data set 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 data sets 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 subspecifications, as well as for the benchmark model and the values of the Akaike information criterion (AIC) and Bayes information criterion (BIC), with both estimation data sets. Inspection of these values, when we include VIX options, suggests that the full SVJ3 specification is substantially superior to all nested subspecifications examined. When VIX options are not part of the estimation data set, the SVJ3 model only slightly outperforms the SVJ2 model, where the intensity of jumps does not load on the additional factor

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

Table 3
RMSRE on the implied volatilities of S&P 500 options (Panel A) and VIX options (Panel B) using the SVJ3 model dynamics given in Eqs. (1)–(3) with different estimation data sets. The columns "w/o" use S&P 500 and VIX index levels and S&P 500 options; "w/" additionally use VIX options. 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.

Panel A: RMSRE on implied volatilities of S&P 500 options	In-sample			In-sample until Sep-08			Out-of-sample		
	w/o	w/	#	w/o	w/	#	w/o	w/	#
Overall	0.096	0.099	15952	0.073	0.079	13077	0.146	0.122	76,540
$Mon. < 0.7$	0.154	0.125	1776	0.079	0.049	1147	0.962	0.963	15,651
$0.7 \le Mon. \le 0.95$	0.091	0.068	6969	0.080	0.055	6098	0.136	0.107	31,703
$0.95 \le Mon. \le 1.05$	0.079	0.095	3848	0.076	0.094	3497	0.097	0.093	15,646
$1.05 \le Mon. \le 1.2$	0.019	0.019	2557	0.017	0.017	2048	0.136	0.126	11,062
$Mon. > 1.2$	0.110	0.116	802	0.050	0.092	287	0.118	0.139	2478
$TTM \le 2\,M$	0.105	0.106	4997	0.085	0.092	4087	0.134	0.113	27,615
$2\,M \le TTM < 6\,M$	0.088	0.078	6069	0.068	0.061	5004	0.144	0.116	28,618
$TTM > 6\,M$	0.106	0.083	4886	0.085	0.062	3986	0.164	0.142	20,307

Panel B: RMSRE on implied volatilities of VIX options	In-sample			In-sample until Sep-08			Out-of-sample		
	w/o	w/	#	w/o	w/	#	w/o	w/	#
Overall	0.611	0.488	4768	0.555	0.372	3562	0.555	0.475	39,612
$Mon. > 1.3$	0.364	0.286	1312	0.095	0.056	4680	0.408	0.322	18,952
$1.1 \le Mon. \le 1.3$	0.386	0.272	1055	0.105	0.057	3120	0.426	0.314	6341
$0.9 \le Mon. \le 1.1$	0.567	0.380	1205	0.507	0.236	965	0.573	0.448	7325
$Mon. < 0.9$	0.942	0.806	1196	0.917	0.691	802	0.887	0.835	6994
$TTM \le 2\,M$	0.624	0.526	2283	0.561	0.403	1662	0.536	0.440	6994
$TTM > 2\,M$	0.582	0.415	2485	0.519	0.289	1900	0.563	0.489	11,355

u. Given that this factor controls for simultaneous jumps in the S&P 500 and VIX index, VIX options contain information that helps identify the parameters driving its dynamics. Therefore, the difference in performance between the SVJ2 and SVJ3 models is larger when VIX options are included in the estimation. Other subspecifications restricting m to a constant, or without jumps, significantly underperform the SVJ2 and SVJ3 models, irrespective of whether or not we include VIX options in the data. The benchmark model, in which positive jumps in the returns are modeled together with negative jumps using a normal distribution, also underperforms the SVJ3 model slightly when excluding and substantially when including VIX options.

Table 3 presents the point estimates and standard errors resulting from the estimation of the SVJ3 model to data sets when either excluding or including VIX options. The estimated parameters driving the two variance processes allow identifying very different roles. Indeed, v has a high volatility parameter σ_v ranging from 0.65 to 0.69 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 0.10, regardless of the estimation data set and chosen model, also with small standard error. Its speed of mean reversion is difficult to estimate precisely but ranges between 0.13 and 0.24, leading to a half-life to three to seven years depending on the estimation data set. 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 the jump process u, σ_u, is close to 0.30, which suggests that u is not as volatile as v but also not as persistent as m. Its 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 data sets.

Similarly, Table 4 presents the point estimates and standard errors resulting from the estimation of the benchmark model to data sets that either exclude or include VIX options.

Tables 5 and 6 report the RMSREs of the model-implied volatilities using, respectively, the SVJ3 and the benchmark model. All statistics are given for both S&P 500 and VIX options, in and out of sample, using the data sets with or

[11] Indeed, under the assumption that jumps have a minor impact on this expectation compared to the drift term, i.e., $\kappa_v^{\mathbb{P}} \gg \lambda_k^{(+/-)} v_v^{\mathbb{P}}$ (inequalities satisfied by our parameter estimates), the conditional expectation of the variance $\mathbb{E}_t^{\mathbb{P}}[v_T]$ can be written as

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

for a constant $c = \frac{\kappa_v}{\kappa_v^{\mathbb{P}} - \kappa_m^{\mathbb{P}}}$. As $\kappa_v^{\mathbb{P}} >> \kappa_m^{\mathbb{P}}$, 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.

Table 4
RMSRE on the implied volatilities of S&P 500 options (Panel A) and VIX options (Panel B) using the benchmark model dynamics given in Eq. (11) with different estimation data sets. The columns "w/o" use S&P 500 and VIX index levels and S&P 500 options; "w/" additionally use VIX options. 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.

Panel A: RMSRE on implied volatilities of S&P 500 options

	In-sample			In-sample until Sep-08			Out-of-sample		
	w/o	w/	#	w/o	w/	#	w/o	w/	#
Overall	0.112	0.117	15,952	0.086	0.089	13,077	0.143	0.119	76,540
$Mon. < 0.7$	0.167	0.142	1776	0.080	0.046	1147	0.962	0.962	15,651
$0.7 \leq Mon. \leq 0.95$	0.094	0.077	6969	0.073	0.053	6098	0.130	0.107	31,703
$0.95 \leq Mon. \leq 1.05$	0.112	0.137	3848	0.106	0.136	3497	0.108	0.102	15,646
$1.05 \leq Mon. \leq 1.2$	0.016	0.016	2557	0.013	0.013	2048	0.085	0.079	11,062
$Mon. > 1.2$	0.167	0.154	802	0.050	0.070	287	0.114	0.112	2478
$TTM \leq 2\,M$	0.099	0.109	4997	0.075	0.095	4087	0.103	0.103	27,615
$2\,M \leq TTM < 6\,M$	0.103	0.100	6069	0.083	0.087	5004	0.145	0.115	28,618
$TTM > 6\,M$	0.146	0.129	4886	0.099	0.084	3986	0.183	0.144	20,307

Panel B: RMSRE on implied volatilities of VIX options

	In-sample			In-sample until Sep-08			Out-of-sample		
	w/o	w/	#	w/o	w/	#	w/o	w/	#
Overall	0.596	0.517	4768	0.533	0.422	3562	0.590	0.496	39,612
$Mon. > 1.3$	0.348	0.279	1312	0.086	0.048	4680	0.402	0.309	18,952
$1.1 \leq Mon. \leq 1.3$	0.362	0.268	1055	0.091	0.050	3120	0.443	0.318	6341
$0.9 \leq Mon. \leq 1.1$	0.527	0.383	1205	0.440	0.221	965	0.633	0.468	7325
$Mon. < 0.9$	0.943	0.877	1196	0.925	0.825	802	0.969	0.902	6994
$TTM \leq 2\,M$	0.584	0.518	2283	0.509	0.407	1662	0.533	0.425	6994
$TTM > 2\,M$	0.586	0.432	2485	0.500	0.278	1900	0.612	0.522	11,355

without VIX options. Options are sorted into buckets by moneyness and maturity.

The pricing errors of the SVJ3 model are, most of the time, lower than those of the benchmark model. In particular, the SVJ3 model provides a better fit to deep out-of-the-money call options on the S&P 500 (RMSRE around 16% for the benchmark model against 11% for the SVJ3 model) and to long-maturity options with horizon above two months. In fact, in the benchmark model, RMSREs increase with time to maturity, from 10% for options with maturity below two months up to close to 15% for when the maturity is larger than six months, when excluding VIX options (from 11 to 13% when including VIX options). With the SVJ3 model, the RMSRE remains around 10%, irrespective of the time to maturity.

However, there are some cases when the benchmark model outperforms the SVJ3 model. For example, this is the case for deep out-of-the-money call options on the S&P 500 out of sample, with an RMSRE of 11% for the benchmark model against close to 14% for the SVJ3 model. Therefore, we will conduct all the coming tests and study the economic implications using both models to ensure our results are robust to the chosen model specification.

5.2. *Is the information in VIX options spanned by S&P 500 options?*

Comparing RMSREs of S&P 500 options in the two scenarios where (i) we exclude VIX options from the estimation data set and (ii) we include them reveals that values are only marginally smaller when they are excluded (0.096 versus 0.099 over all S&P 500 options in the in-sample period). Constraining the model to fit VIX option prices therefore does not significantly affect the fit to S&P 500 option prices. 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, RMSEs[12] even decrease from 0.044 to 0.037 when we do not include VIX options in the estimation.[13] The improvement is most noticeable for deep OTM put options (RMSEs decrease from 0.102 to 0.086), OTM put options (0.027 versus 0.034), and long-term options (0.032 versus 0.039), suggesting that VIX options help to improve the identification of model parameters driving the prices of these options. Intuitively, VIX options help pinpoint stark volatility increases 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.

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

[13] Our results are comparable to those obtained by Andersen et al. (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 1.3%, and 3.7% when including the crisis period. In addition, their model has an additional volatility factor hence adding flexibility to fit more data.

Table 5

Estimated parameters of the following regression: $\text{returns}(t, t+h) = \alpha + \beta * \{\text{set of predictors}\}_t + \epsilon_t$. The first two sets of predictors include the past RV, the model-implied one-month VRP (mVRP) and the slope of its term structure (VRP TS), measured as the three-month VRP minus the one-month VRP. Results are displayed for the benchmark model (bm) and for the SVJ3 model dynamics given respectively in Eqs. (1)–(3) and (11). The third set of predictors includes the past realized variance and the model-implied VRP only. The fourth set of predictors include the latent factors of the SVJ3 model. Finally, the last predictor is a model-free estimate of VRP, i.e., the difference between the realized variance over the past 30 days and the current VIX value squared (RV-VIX^2). Below each estimate, we report its Hansen–Hodrick adjusted standard error.

	h = 1 month					h = 3 months					h = 5 months				
Model	bm	SVJ3	SVJ3	SVJ3	Model-free	bm	SVJ3	SVJ3	SVJ3	Model-free	bm	SVJ3	SVJ3	SVJ3	Model-free
constant	−0.035	*0.001*	*−0.009*	−0.050	−0.009	−0.076	0.004	−0.019	0.143	−0.020	−0.107	−0.001	−0.030	−0.130	−0.023
	0.007	*0.006*	*0.007*	*0.041*	*0.007*	*0.030*	*0.022*	*0.023*	*0.102*	*0.023*	*0.058*	*0.038*	*0.042*	*0.241*	*0.036*
v				0.951					2.201					2.873	
				0.108					*0.537*					*0.878*	
m				−0.281					−0.662					−0.706	
				0.102					*0.575*					*0.899*	
u				0.059					−0.120					0.154	
				0.037					*0.113*					*0.232*	
RV	−1.005	−1.009	−0.722	−1.005		−2.185	−2.172	−1.538	−2.185		−2.878	−2.888	−2.116	−2.878	
	0.090	*0.090*	*0.126*	*0.090*		*0.500*	*0.505*	*0.304*	*0.510*		*0.908*	*0.916*	*0.518*	*0.884*	
mVRP	−7.961	−12.636	−6.078			−17.779	−27.792	−13.274			−23.373	−37.167	−19.364		
	0.695	*1.134*	*1.241*			*5.821*	*9.023*	*3.139*			*9.926*	*15.406*	*5.471*		
VRP TS	12.996	9.161				29.072	20.354				37.925	25.078			
	1.145	*1.590*				*10.143*	*12.241*				*17.090*	*19.105*			
RV-VIX^2					−0.772					−1.841					−2.444
					0.120					*0.553*					*0.736*
Adj R^2	0.122	0.123	0.101	0.126	0.097	0.169	0.167	0.135	0.174	0.167	0.155	0.156	0.131	0.160	0.151
	h = 6 months					h = 7 months					h = 8 months				
Model	bm	SVJ3	SVJ3	SVJ3	Model-free	bm	SVJ3	SVJ3	SVJ3	Model-free	bm	SVJ3	SVJ3	SVJ3	Model-free
constant	−0.128	0.004	−0.036	−0.008	−0.022	−0.136	0.007	−0.037	0.045	−0.019	−0.131	0.008	−0.035	−0.071	−0.013
	0.070	*0.043*	*0.052*	*0.258*	*0.040*	*0.085*	*0.047*	*0.061*	*0.254*	*0.047*	*0.094*	*0.053*	*0.068*	*0.271*	*0.052*
v				3.486					3.769					3.637	
				1.029					*1.276*					*1.262*	
m				−1.128					−1.299					−1.238	
				0.999					*1.166*					*1.179*	
u				0.044					−0.004					0.112	
				0.262					*0.261*					*0.269*	
RV	−3.250	−3.250	−2.186	-3.250		−3.430	−3.425	−2.243	−3.430		−3.286	−3.290	−2.139	−3.286	
	1.017	*1.030*	*0.592*	*1.002*		*1.220*	*1.233*	*0.670*	*1.210*		*1.220*	*1.233*	*0.698*	*1.194*	
mVRP	−29.127	−45.878	-21.350			−31.6928	−49.780	−22.474			−30.862	−48.669	−22.101		
	11.136	*17.422*	*6.625*			*13.552*	*21.163*	*7.401*			*13.672*	*21.371*	*7.919*		
VRP TS	47.728	34.581				52.066	38.571				50.666	37.494			
	19.120	*21.434*				*23.244*	*25.503*				*23.428*	*25.761*			
RV-VIX^2					−2.648					−2.675					−2.501
					0.735					*0.835*					*0.778*
Adj R^2	0.152	0.152	0.116	0.154	0.138	0.146	0.146	0.108	0.147	0.118	0.121	0.122	0.091	0.123	0.090

Table 6
Descriptive statistics for daily S&P 500 futures returns and daily VIX returns and square levels for the periods from March 2006 to February 2009 and March 2009 to April 2016. We report the mean (Mean), standard deviation (Std), skewness (Skew), and kurtosis (Kurt).

	March 2006–February 2009				March 2009–April 2016			
	Mean	Std	Skew	Kurt	Mean	Std	Skew	Kurt
S&P 500 ret.	−0.0009	0.0170	−0.1487	11.5442	0.0006	0.0107	−0.0879	7.4032
VIX ret.	0.0048	0.0774	1.3553	10.7419	0.0020	0.0750	1.2731	8.1333
VIX^2	0.0730	0.1038	2.7325	10.7644	0.0427	0.0341	2.2688	8.8821

Much more interesting is the comparison of pricing errors on VIX options using our different data sets. VIX options are not well priced when they are not in the estimation data set, i.e., the information they contain is not spanned by S&P 500 options. This is true in and out of sample, outside and during the crisis period. In sample, the RMSRE over all VIX options decreases from 0.611 to 0.488 when we add VIX options to the estimation data set. Such a decrease is observed throughout all maturity and moneyness buckets. The corresponding RMSE goes down from 0.402 to 0.276. Out of sample, the numbers are comparable, with a decrease of the RMSRE from 0.555 to 0.475 and a decrease of the RMSE from 0.401 to 0.350. We note that the out-of-sample pricing errors are smaller than the in-sample errors, which is due to the fact that the financial crisis overlaps to a large part with our in-sample period.

The mispricing of VIX options, when they are not included in the estimation data set, indicates that the conditional $\mathbb{Q}$ -distributions of variance over the time horizons covered by options' maturities are not well represented. A comparison of the gain in RMSREs between the VIX options expiring in less than two months and the others, when including VIX options in the data set, confirms this result. In sample (out of sample), the gain for short-term options is indeed 0.099 (0.096), versus 0.167 (0.074) for long-term options. Results are unchanged when using the benchmark model.[14]

To assess whether the differences between pricing errors obtained with the two estimation data sets are significant, we run Diebold–Mariano tests. The time-t loss function is given by the mean square relative error between model-implied option prices and observed prices. Denote the loss differential between the errors produced with estimation data sets without and with VIX options by d_t. Under the null hypothesis that the two estimation data sets produce pricing errors of equal magnitude, $\mathbb{E}[d_t]=0$. If the estimation including VIX options produces smaller pricing errors than the estimation excluding them, then $\mathbb{E}[d_t]>0$. The Diebold–Mariano statistic is the $t-$statistic for this test. Table 7 reports the results.[15] The Diebold–Mariano tests strongly confirm that VIX options are better priced when included in the estimation data set. Additional model diagnostics are reported in the online Internet Appendix, Section C.

Table 7
Log-likelihood (LL) and values of the AIC and BIC tests for subspecifications of the SVJ3 model dynamics given in Eqs. (1)–(3) and for the benchmark model (dynamics given in Eq. (11)). The specifications considered are no jumps (SV2), m and u are constant (SVJ), u is constant (SVJ2), and no negative jumps in the variance ($SVJ3^{\Delta v>0}$). The benchmark model has two factors and normally distributed jumps in the returns and their variance. The estimation period is from March 2006 to February 2009.

	Without VIX options					
	SV2	SVJ	SVJ2	$SVJ3^{\Delta v>0}$	Benchmark	SVJ3
LL	9778	10,224	10,424	10,415	10,370	10,473
AIC	−19,538	−20,414	−20,802	−20,784	−20,704	−20,898
BIC	−19,496	−20,335	−20,695	−20,677	−20,621	−20,787
			With VIX options			
			SVJ2	$SVJ3^{\Delta v>0}$	Benchmark	SVJ3
LL			9717	9729	9538	9949
AIC			−19,388	−19,412	−19,040	−19,850
BIC			−19,281	−19,305	−18,957	−19,739

5.3. Filtered trajectories

In Fig. 2, we plot the trajectories of the volatility processes v and m, filtered using the SVJ3 and the benchmark models. Panel A represents the trajectory of v including VIX options in the estimation data set. The trajectories for v obtained using the SVJ3 and the benchmark model overlap almost perfectly. Note that v never touches or crosses the zero boundary.

Panel B represents the filtered trajectory of the stochastic central tendency m. 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 to a level, in 2016, that is slightly above the initial level of 2006. The high volatility due to the flash crash in May 2010 translates into an slight increase of m from 5 to 7%. While the process v reaches

[14] We also analyze pricing errors for subspecifications of the SVJ3 model. They are larger due to the restrictions imposed but confirm the observations made for the SVJ3 model. In unreported results, we find that the stochastic central tendency significantly improves the pricing of long-term options and the representation of the tails of the distributions of 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. Pricing errors for subspecifications of the SVJ3 model are available upon request.

[15] The standard errors are computed using the Newey and West (1987) autocorrelation and heteroskedasticity consistent variance estimator with the number of lags optimally chosen according to Andrews (1991).

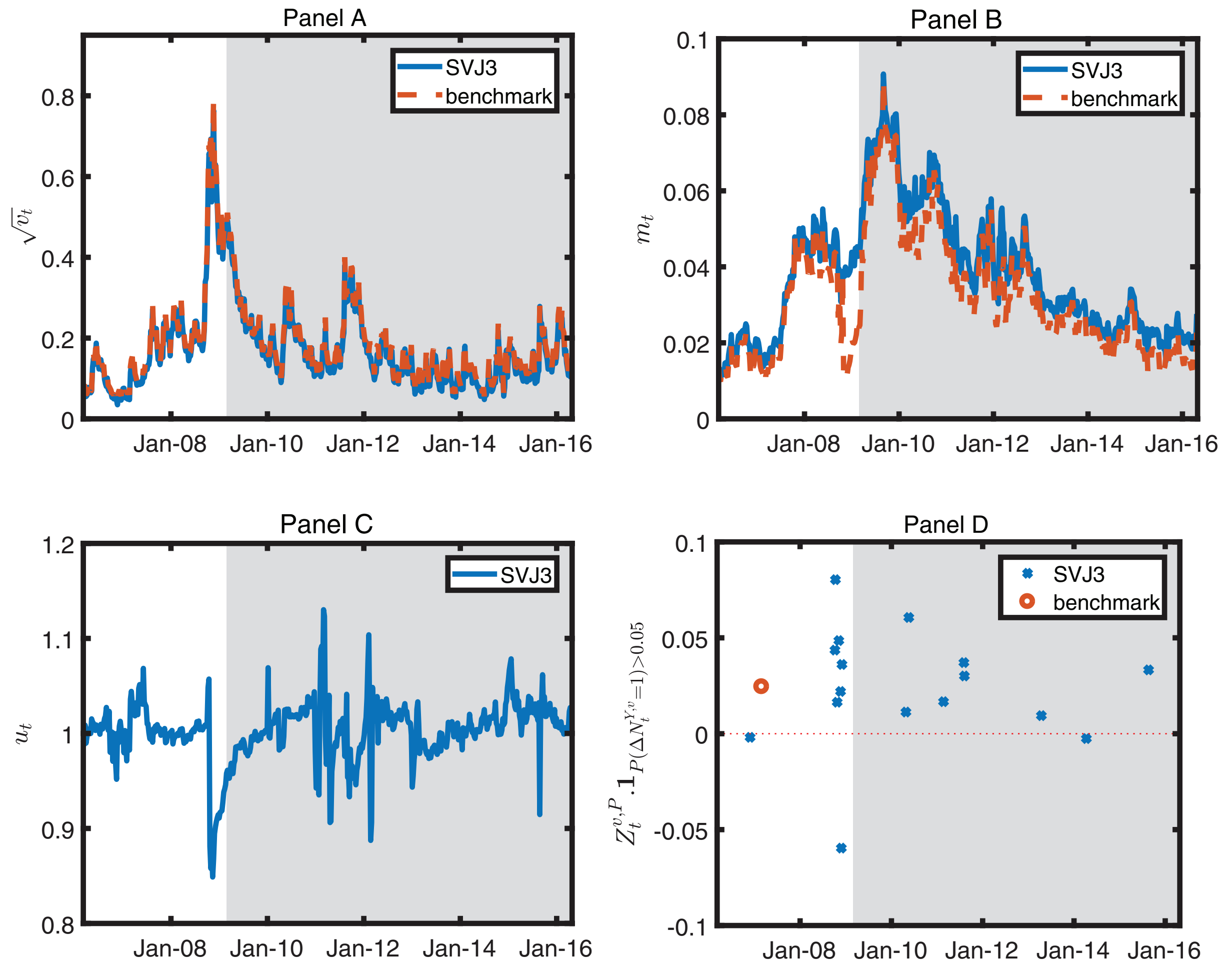

Fig. 2. Panels A and B plot the filtered trajectories of the latent volatility processes v and m for the SVJ3 and the benchmark model given in Eqs. (1)–(3) and (11), respectively, including VIX options in the estimation data set, from March 2006 to April 30, 2016. Panel C represents the filtered trajectory of the jump intensity factor u, and Panel D displays the filtered size of jumps in the variance process v, when the probability of jumps exceeds 5%. The shaded part of the graph represents the out-of-sample period, from March 1, 2009, until the end of April 2016.

levels that are close to its initial level from the beginning of 2010, m reverts at a much slower pace.

In Panel C, we plot the filtered trajectory of the jump intensity factor u. The factor u acts as correction factor in times of market stress and wildly oscillates about its long-term mean $\theta_u = 1$. The factor u therefore corrects for the nonlinearity in v of the intensity of jumps. Panel D represents the recovered jump sizes in v, when the estimated probability of jumps is larger than 5%. Note that few negative jumps in the volatility are filtered, the main one occurring at the peak of volatility, hence causing no risk that the volatility level crosses zero.

5.4. Representation of the variance term structure

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 that are not included in the estimation data set. 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 three-month VXV index and in 2013 the nine day VXST index. We build indices for maturities of two, three, five, and six months and verify that our model can reproduce them. We calculate the RMSEs for the SVJ3 and the benchmark model. The RMSE on the VIX square during the in-sample period, obtained from daily data, is 2.6%, and goes down to 1.6% when excluding the data after September 2008 from the calculation. It is equal to 1.8% in the out-of-sample period. The VIX index is therefore well fitted, whether VIX options are in the estimation data set or not. 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 six-month VIX to be less precisely synthesized than the 30-day VIX index. RMSEs are reported in Table 8.

Including VIX options in the estimation data set, we obtain RMSEs on the synthesized two-, three-, five-, and six-month squared VIX indexes, obtained from weekly data,

Table 8
Statistics of Diebold–Mariano tests applied to option pricing errors. The pricing error at time t is defined as the mean square relative error between the model-implied option prices and observed prices. The test statistics compare the errors when excluding VIX options to the case when including VIX options. Under the null hypothesis that the two estimation data sets produce pricing errors of the same magnitude, the Diebold–Mariano test statistic is standard normal. A positive value means that the pricing errors become smaller when including VIX options.

	In-sample	Out-of-sample
S&P 500 options		
Overall	0.908	9.488
$Mon. < 0.7$	0.273	−6.252
$0.7 \leq Mon. \leq 0.95$	0.719	3.327
$0.95 \leq Mon. \leq 1.05$	−0.158	0.400
$1.05 \leq Mon. \leq 1.2$	0.000	0.693
$Mon. > 1.2$	−0.181	−0.167
$TTM \leq 2\,M$	−0.063	3.535
$2\,M \leq TTM < 6\,M$	0.338	2.712
$TTM > 6\,M$	0.422	5.721
VIX options		
Overall	111.753	133.872
$Mon. > 1.3$	7.432	26.020
$1.1 \leq Mon. \leq 1.3$	3.047	6.976
$0.9 \leq Mon. \leq 1.1$	3.130	15.870
$Mon. < 0.9$	17.423	132.233
$TTM \leq 2\,M$	13.062	23.411
$TTM > 2\,M$	9.928	120.845

between 1.4 and 2.5% over the entire time period considered (in and out of sample). In contrast, when we exclude VIX options from the estimation data set, we obtain RMSEs between 1.5 and 3.1%. These results provide striking evidence that VIX options play a fundamental role in identifying the parameters of the model that drive the variance term structure. The same ordering is noticed for both models examined, namely the SVJ3 and the benchmark model. This result holds particularly true in the in-sample period, which is much more hectic than the out-of-sample period. Excluding the period starting in September 2008, i.e., the market distress period, from the in-sample period, does not change the result. Therefore, VIX options provide information over the whole time period, disregarding market conditions. This holds for both models examined.

6. Variance risk premium

From Section 5 we learn that including VIX options in the estimation data set allows us to achieve a more precise representation of the risk-neutral conditional distribution of future variance. The VRP in Eq. (13) is, by construction, highly dependent on the dynamics of variance under both historical and risk-neutral measure. In this section, we analyze the dynamic properties of the VRP, using the estimation performed with the full data set, i.e., we include VIX options.

6.1. Estimated VRP

We estimate the VRP for investments with different times to maturity, varying from one week to one year. In Fig. 3, Panel A, we plot the evolution of the VRP over time, with its 90% confidence interval. In line with literature, we find that the VRP is negative. At the beginning of our sample period, it is 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 six-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 2010 flash crash 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 of Fig. 3 represents the VRP for different times to maturity, ranging from one month to one year. The sign and shape of the VRP are consistent across maturities. Before September 2008, the VRP range between 0 and -2.5%, exhibiting a flat term structure. 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 Fig. 3, we decompose the one-week and six-month VRP into their continuous and discontinuous parts. The discontinuous component of the VRP dominates for shorter maturities, indicating that including jumps helps the representation of the shorter end of the VRP's term structure. In fact, the jump part of the VRP constitutes about two-thirds of the total VRP for a one-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 six-month VRP.

Finally, Panels E and F of Fig. 3 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 one-week VRP recovers much faster than the six-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. Hence, summarizing the above dicussion, we find that 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.

Fig. 4 compares the VRP to its model-free ex-ante estimate and to its ex-post approximated value, for the one-month and six-month horizons. The ex-ante estimate is computed as the difference between the past realized variance (over the last 30 days for the one-month VRP and over the last six months for the six-month VRP) and the current value of the VIX index. As the expectation of the future realized variance is not available in a model-free way, this estimate assumes that it can be approximated by the past realized variance. The ex-post approximation is calculated as the difference between the observed realized

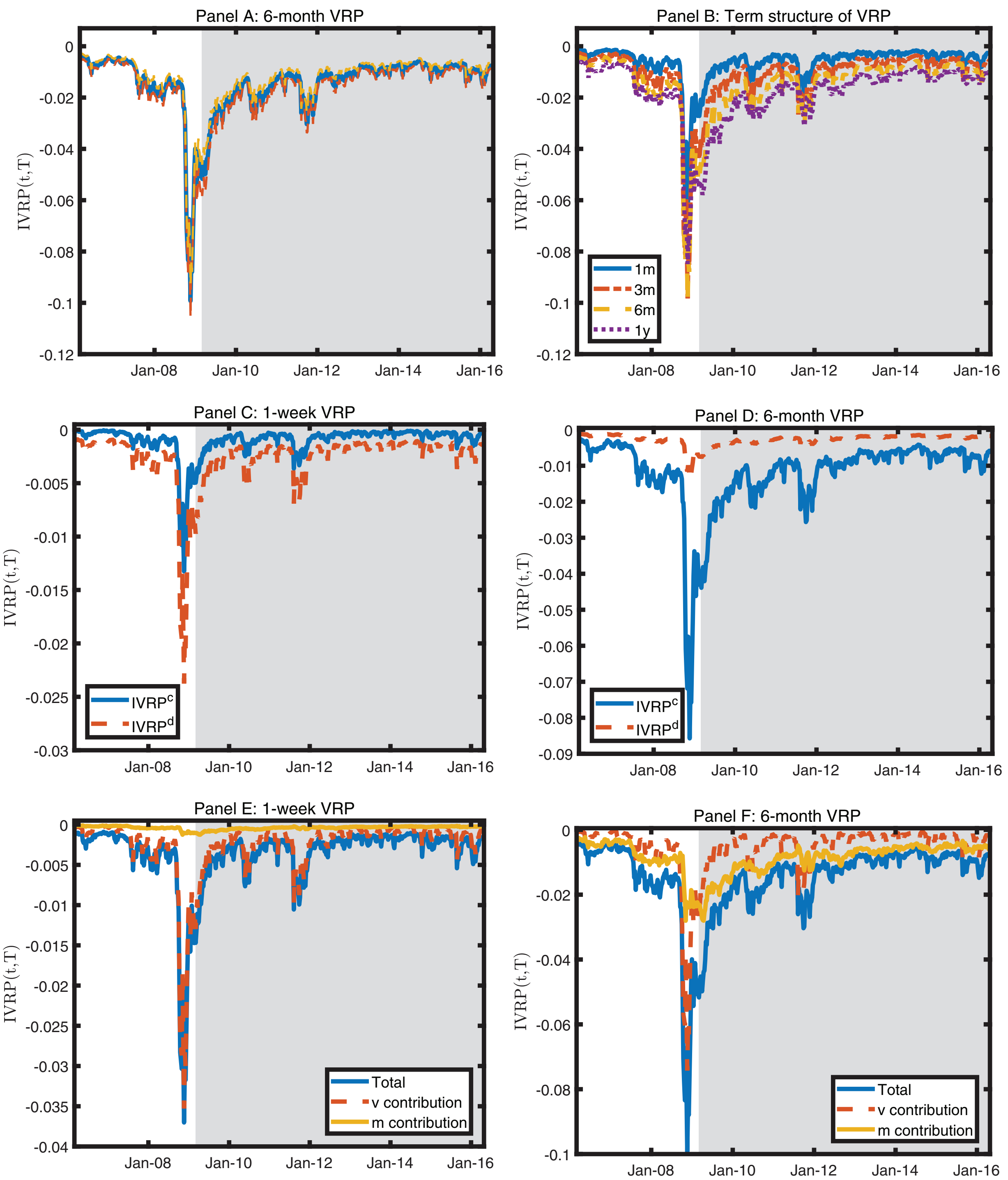

Fig. 3. Integrated VRO when estimating the SVJ3 model dynamics given in Eqs. (1)–(3) using the full data set of S&P 500 options, VIX options, and their underlying levels. Panel A plots the six-month VRP and its 90% confidence interval, conditional on parameter estimates. Panel B plots the VRP for different maturities. Panels C and D decompose the one-week and six-month VRP into their continuous and discontinuous components. Panels E and F plot the contribution of the latent factors v and m to the one-week and six-month VRP. The shaded parts of the graphs represent the out-of-sample period from March 1, 2009, until the end of April 2016.

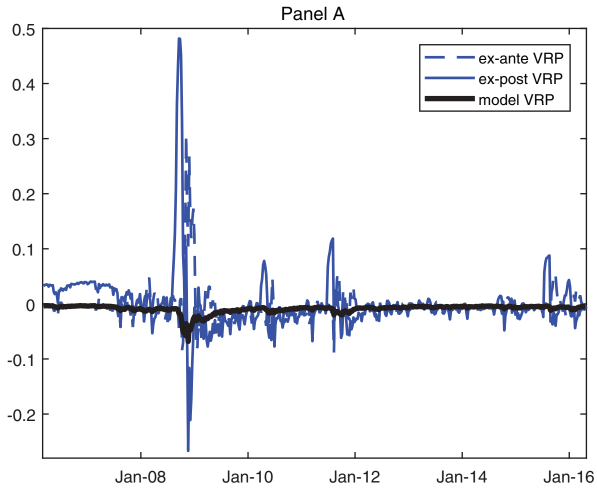

Fig. 4. Integrated VRP when estimating the SVJ3 model dynamics given in Eqs. (1)–(3) using the full data set including VIX options (thick solid line), compared to the ex-ante model-free estimate of the VRP, computed as the difference between the past realized variance and the squared VIX (dashed line), and the approximated ex-post VRP, computed as the difference between the realized variance and the squared VIX (thin solid line). Panel A corresponds to an investment of one month and Panel B to six months.

variance at maturity and the VIX value at the beginning of the observation period. In times of market turmoil, the realized variance is much higher than its expected value, causing the ex-post VRP to be highly positive and to vary in an erratic way around it conditional first moment.

6.2. 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 recently. Indeed, in contrast to most papers, finding a VRP that is more negative when time to maturity increases (Amengual, 2008; Gruber et al., 2015, among others) instead find 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 by van Binsbergen et al. (2013) and is intuitively appealing from an economic perspective. Indeed, in low volatile periods, investors are likely to require a larger compensation for a long-term investment, as the probability that markets enter into distress before expiry increases with time horizon. However, 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 expiry of the investment. This increase in probability justifies a lower (in absolute value) VRP for long-term investments during of market distress.

Panel A of Fig. 5 plots the difference between the three-month and the nine-month VRP over time. This difference is generally positive, meaning that the 9-month VRP is more negative than the three-month VRP. During the recession (shaded area), this relation is inversed, which is in line with the findings of Gruber et al. (2015), and the difference becomes highly negative. In terms of modeling, 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 Fig. 5 represent the continuous and jump contributions in our estimation and show that, even though both contributions of the VRP to the three-month minus one-year difference are responsible for the change in sign, the jump component switches more often than the continuous component.

As the above results could be the artefact of our model specification, we seek to confirm them in a model-free way. The $\mathbb{P}$ -leg of the VRP is not observable. However, the $\mathbb{Q}$ -leg can be well approximated, up to an adjustment due to jumps, by the VIX index. In Fig. 5, Panel D plots the difference between the three-month and the six-month VIX, constructed from S&P 500 options following the procedure used in Section 5.4. It indicates that the slope of the term structure of the $\mathbb{Q}$ -expectation of total variance becomes strongly positive during the recession, driving the slope of the term structure of the variance risk premium down. This observation suggests that our results are not driven by model misspecification.

Another property of the VRP term structure that has been recently investigated in the literature is its convexity (or concavity, in papers that define the VRP as the $\mathbb{Q}$ -expectation of variance minus the $\mathbb{P}$ -expectation). See, e.g., Andries et al. (2015) and Dew-Becker et al. (2017). 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 flattens out for longer term maturities. In Panel A of Fig. 6, we plot the term structure of VRP on different dates and show 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 flat after the three-month horizon, corroborating the findings of Dew-Becker et al. (2017). However, in October 2008, the shape of the VRP term structure radically changes. It is no longer flat

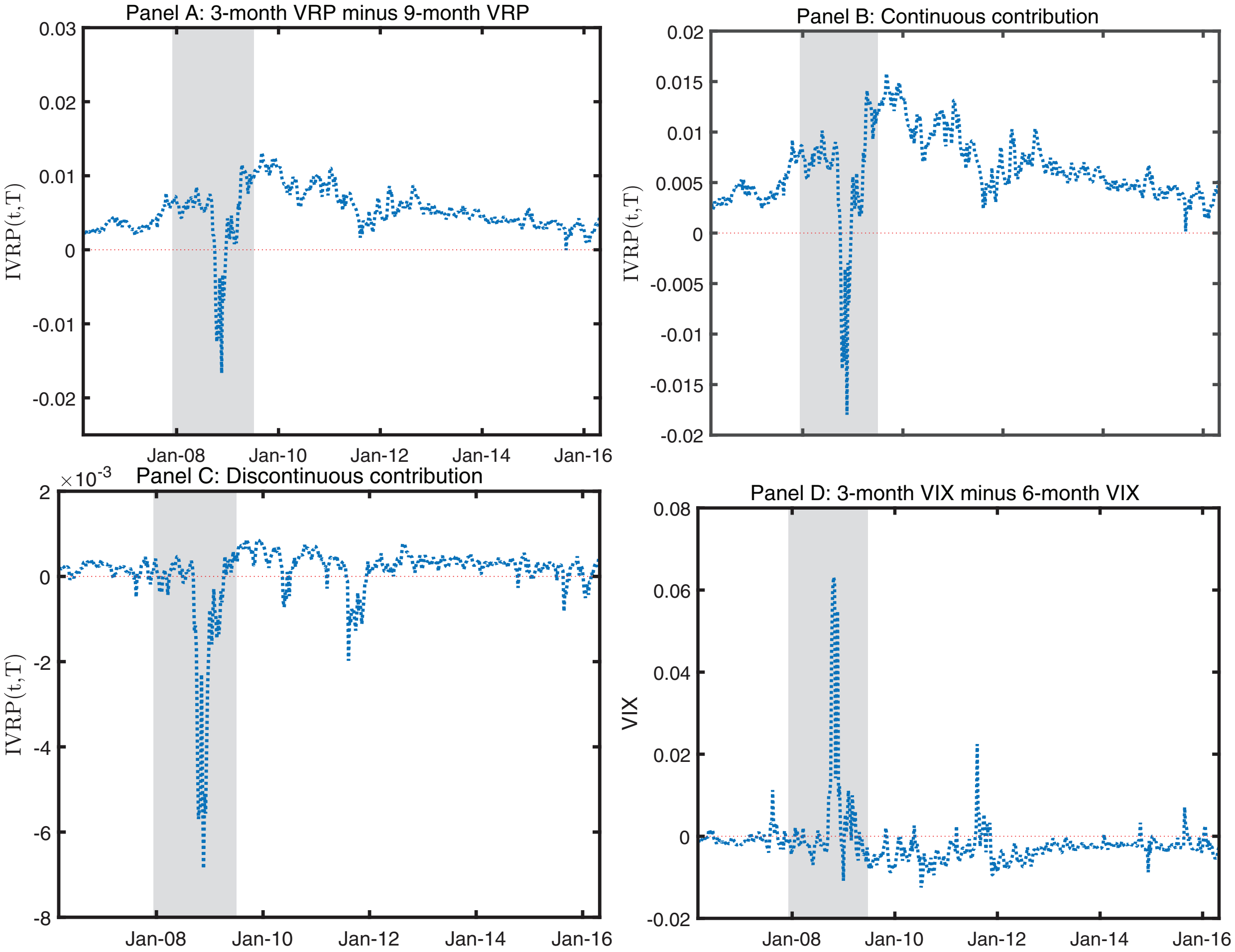

Fig. 5. Panel A displays the total difference between the three-month and the 9-month integrated VRP when estimating the SVJ3 model dynamics given in Eqs. (1)–(3) using the full data set of S&P 500 options, VIX options, and their underlying levels. Panel B plots the contribution of the continuous fluctuations in the VRP to this difference and Panel C the contribution of the jumps. Panel D plots the difference between the three-month and the six-month VIX constructed from S&P 500 options. The shaded part of the graphs represents the NBER recession from December 2007 until June 2009.

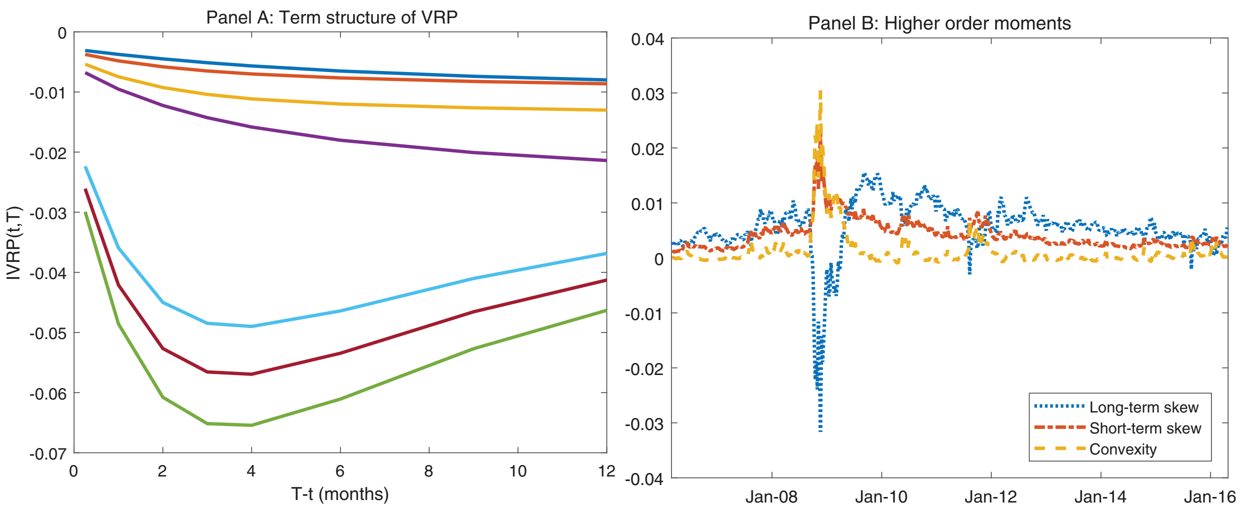

Fig. 6. Panel A represents the term structure of variance risk premium for investment horizons from one month to one year. The VRP is calculated on the following dates, going from the upper curve to the lower one: March 3, 2006, May 17, 2006, June 14, 2006, June 18, 2008, November 5, 2008, December 10, 2008, October 22, 2008. Panel B plots the long- and short-term skewness and the convexity of the term structure of VRP over time.

for long horizons, and, moreover, it exhibits a sharp decline up to a horizon of around three 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 market's belief that the situation could get even worse in the short term, but will eventually get better in the long term.

Panel B of Fig. 6 represents an estimate of the higher order moments of the VRP term structure. The short-term skewness is measured as the difference between the one-month and the three-month VRP, while the long-term skewness represents the three-month VRP minus the one-year VRP. They hence represent the insurance premium an investor would be willing to pay to be hedged against fluctuations in the forward variance. The convexity is measured as the sum of the one-month and the six-month VRP minus two times the three-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 coefficient above 0.96.

Our results are complementary to the ones of Gruber et al. (2015) and Dew-Becker et al. (2017). Compared to the former, we only find a switch in the slope of the VRP term structure over mid- to long-term maturities. For short-term horizons, we find that the slope becomes more pronounced in times of market turmoil, causing our convexity indicator to increase. Compared to the latter, we find that investors do care about shocks in future variance but that the way they care differs across time, depending on economic conditions.

7. Economic implications

In this section, we analyze two important applications of our model. First, we seek to use the information contained in the VRP term structure as a trading signal for a volatility strategy. Second, we investigate the predictive power of our VRP estimates on future S&P 500 returns.

7.1. Investing in variance

The switch in the slope of the VRP term structure has strong economic implications. Indeed, 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.[16] We start our investment analysis by asking whether the term structure of VRP can be used as trading signal. In the spirit of Dew-Becker et al. (2017), we define a squared VIX forward as a claim on the future variance with the following payoff:

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

Hence, $\mathrm{VIX}^2_{t,T_1,T_2}$ represents the value of a portfolio that is long variance with maturity T_2 and short variance with maturity T_1.[17] Assuming that there exists an index tracking the squared VIX forward on a weekly basis, the annualized Sharpe ratio of an investment in the $\mathrm{VIX}^2_{t,3\text{ months},6\text{ months}}$ from 2006 until 2016 is equal to 0.01. On average, a simple trading strategy that is long forward variance is therefore not profitable for this period, in line with Dew-Becker et al. (2017).[18]

Taking a closer look at the data and, in particular, at the slope of the VRP term structure, we find that this meager performance can be explained as follows. As long as realized variance is low, i.e., outside the financial crisis, the strategy yields a positive payoff equal to the difference between the three-month and six-month VRP. This difference becomes negative during the financial crisis, causing losses for the portfolio holder and canceling out the previous gains.

Therefore, we can interpret a switch in the sign of the slope of the VRP term structure as a warning that the future realized variance may increase, as a result of which the forward variance risk premium is no longer positive. If one leaves the investment on hold until the slope switches sign again, one can avoid some of the losses of the former strategy and generate a Sharpe ratio of 0.46. Switching position from selling future variance into buying future variance whenever the slope of the VRP term structure is negative further enhances the Sharpe ratio, which reaches 0.77.

The returns on forward variance and the times when the strategy is kept on hold are displayed in Fig. 7. The investment in forward variance is only interrupted seven weeks during the whole time period, starting in October 2008. The large improvement in the Sharpe ratio is achieved during this short period of time.

Using the benchmark model, these results still hold, and Sharpe ratios become, respectively, 0.17 and 0.32. Hence, we can interpret the sign of the VRP slope as a trigger, indicating that selling future variance is no longer profitable. These results are robust to other values of T_1 and T_2. For example, the previously reported Sharpe ratios become -0.01, 0.10, and 0.15, respectively, if using $T_1 = $ four months instead of three months and 0.01, 0.21 and 0.23 if using $T_2 = 9$ months. In the latter case, our algorithm detects an additional switch in August 2015, triggered by the Chinese yuan devaluation.

7.2. Return predictability

Bollerslev et al. (2009), using the difference between implied and realized variances as a proxy for the VRP, find

[16] Such investment has been shown to be dynamically optimal in Egloff et al. (2010) and Filipović et al. (2016) using variance swaps.

[17] Dew-Becker et al. (2017) build these claims for $T_2 - T_1 = 1$ month. Such payoff can be attained by a long-short strategy in variance swaps with maturities T_1 and T_2 or equivalently (up to a jump term) by a portfolio of S&P 500 options.

[18] Dew-Becker et al. (2017) note that for T_2 larger than 2 months and $T_2 - T_1 = $ a month, Sharpe ratios are "insignificantly different from zero." The disappointing performance for this simple trading strategy is mostly caused by punctual losses occurring during the financial crisis in 2008.

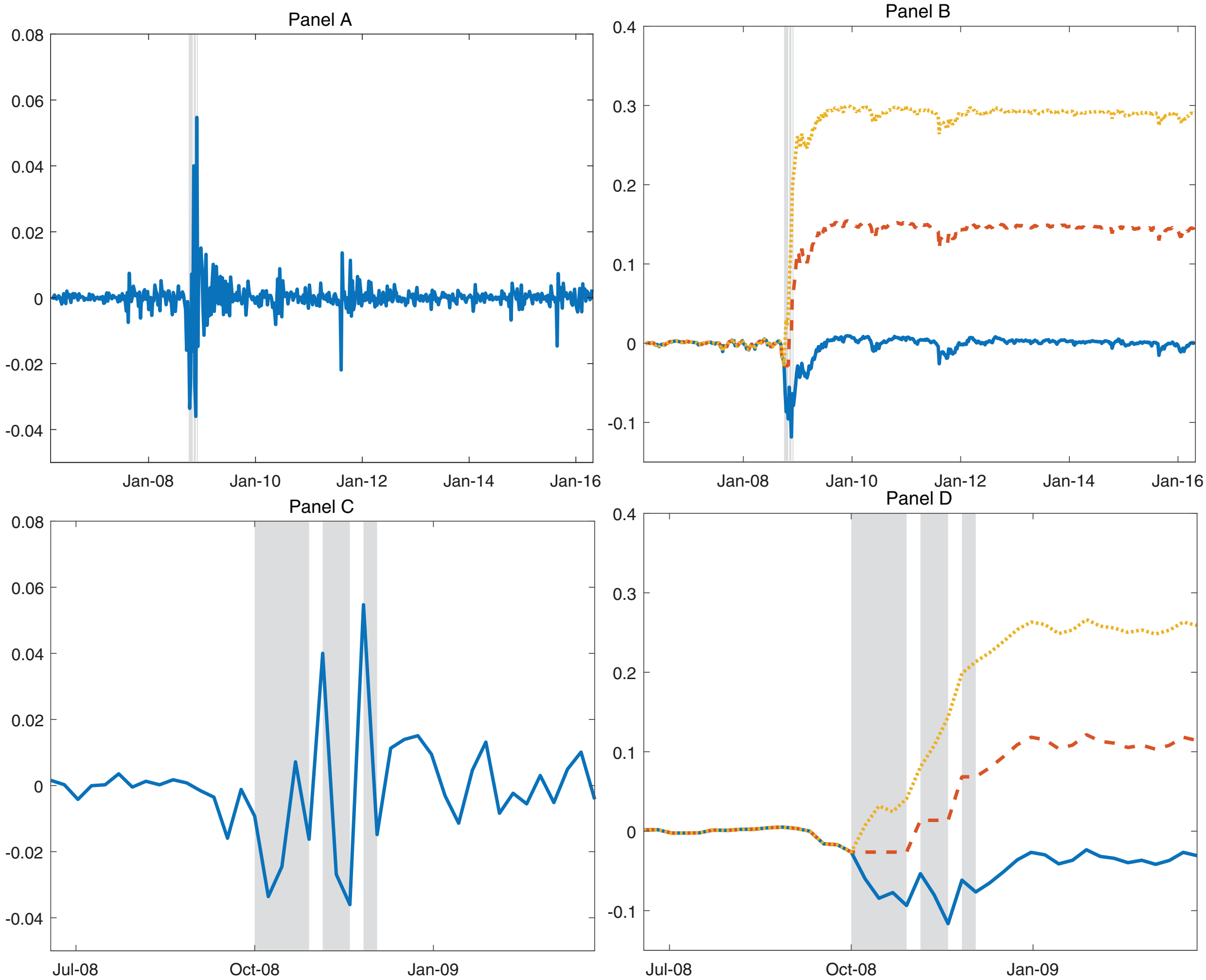

Fig. 7. Panel A represents the returns of a weekly investment in the forward variance, with $T_1 = 3$ months and $T_2 = 6$ months. The gray areas indicate the weeks in which there is a switch in the slope of the VRP term structure. Panel B displays the cumulative returns of this investment and the variants that we propose. The plain line (lower line) corresponds to the usual long-short strategy, the dashed line (middle line) to the strategy that is kept on hold whenever there is a switch in the slope of the VRP term structure, and the dotted line (upped line) to the strategy that takes the opposite position of the usual strategy during these periods of switches. Panels C and D are zooms of Panels A and B, which focus on the period during which such switches are observed.

that it has predictive power on future returns, with large premia predicting high future S&P 500 returns. Their finding is robust and has been confirmed in subsequent papers.[19] More generally, as we work in an affine setup and the VRP is a linear function of the state variables, we investigate the predictive power of the three state variables in our model on S&P 500 returns.

Our results are summarized in Table 9, which reports the estimated parameters for all regressions, using monthly observations.[20] As we see from Table 9, the beta of the VRP is negative, confirming the usual results: the larger the VRP (in absolute value), the larger future returns on average. In contrast, the beta of the term structure variable is always positive, indicating that a larger short-term skewness goes together with higher future returns on average. We also include the difference between the VIX and the (past) one-month realized variance as predictor, as in Bollerslev et al. (2009). As expected, the coefficient for this predictor is negative. The coefficient in front of the short-term variance v is positive. Intuitively, the VRP loads heavily on the v factor, with negative loading. The negative coefficient in front of the VRP therefore translates into a positive coefficient in front of the v process. The coefficient in front of the long-term variance process m is negative, indicating that, on average, the larger the stochastic central tendency, the smaller the future returns. The coefficient in front of the negative jump intensity process u changes with maturity. It is positive for the one-month horizon, with a t-statistic of 1.584. For longer

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

[20] Our results should be interpreted with caution due to the overlapping windows for horizons larger than a month.

Table 9
RMSE on the synthesized squared VIX indices for maturities of two, three, five, and six months, excluding or including VIX options in the estimation data set. Results are reported over the entire time period (Overall), the in-sample period (IS), the in-sample period until September 2008 (IS S08), and the out-of-sample period (OOS). The indices are synthesized from S&P 500 options.

	w/o VIX options				w/ VIX options			
	Overall	IS	IS S08	OOS	Overall	IS	IS S08	OOS
SVJ3 model								
2m	0.018	0.017	0.008	0.018	0.015	0.013	0.007	0.015
3m	0.023	0.022	0.011	0.023	0.019	0.017	0.006	0.019
5m	0.029	0.028	0.015	0.030	0.024	0.021	0.007	0.024
6m	0.031	0.030	0.016	0.031	0.025	0.022	0.008	0.026
Benchmark model								
2m	0.017	0.018	0.016	0.015	0.014	0.014	0.007	0.014
3m	0.020	0.023	0.018	0.018	0.018	0.018	0.007	0.017
5m	0.025	0.028	0.022	0.021	0.022	0.022	0.010	0.022
6m	0.026	0.031	0.024	0.022	0.024	0.024	0.012	0.023

horizons, it loses significance, confirming the intuition that jumps only impact returns on the short term.

Similarly to Bollerslev et al. (2009), the R^2 increases with time to maturity, reaches its maximum around four months, and then goes down. This pattern holds for all regressions. The main difference between using the model-free estimator and the model-implied VRP estimates comes from the rate at which the R^2 declines after reaching its optimum. For a five month maturity, the model-free VRP exhibits an adjusted R^2 of 0.151 against 0.160 for the model predictors, but for a seven month horizon the gap increases: 0.118 versus 0.147; for eight months these numbers become 0.09 versus 0.123. Although the difference is relatively small, this gap suggests that adding model-implied predictors adds value to the predictions. This result holds for both the benchmark model and the SVJ3 model.

Finally, this exercise highlights the information content of the slope of the VRP term structure. Indeed, running the regression only with the past RV and the VRP yields an adjusted R^2 that is substantially lower than the one obtained when adding the slope of the VRP term structure, for all horizons considered. This result can be explained in terms of projections on the latent factors. The VRP, in our models, is explained by two to three latent factors, depending on whether we use the benchmark or the SVJ3 model. The slope of the VRP term structure is described by these factors as well. Adding it to the regression therefore allows spanning the information available in the two dominant factors v and m. For the benchmark model, the spanning is perfect in the sense that the VRP and the slope of its term structure are perfectly described by a weighted average of v and m. For the SVJ3 model, there is a slight loss of information. This loss of information is quantified by the differences between the second and the fourth regressions, i.e., the regression using the VRP and its term structure and the one using all latent factors v, m, and u. The resulting adjusted R^2 are fairly similar, the largest difference being attained for a three-month horizon: 0.167 versus 0.174.

Our results therefore show that similarly to the variance that is described by two to three factors, the VRP contains more than one dimension of interest, including its level but also the slope of its term structure. The latter dimension matters not only for its implications in terms of trading strategy but also for its valuable information content on future returns.

8. 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 data sets. The first data set contains a time series of S&P 500 and VIX indexes, as well as S&P option prices, and the second data set contains VIX options in addition to the first data set. We do not restrict the moneyness and maturity of the options considered so that we can fully benefit from their information content 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 data set, they are not well priced. Including them in the estimation data set not only improves the pricing of VIX options in both in- and out-of-sample periods but also the pricing of 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 data set allows considerable improvements in representing the variance term structure. We replicate the construction of the VIX index for maturities two, three, five, and six months. We find that for maturities larger than three months, the model yields RMSEs which are significantly larger when VIX options are excluded from the data set.

We perform a thorough analysis of the VRP and of its term structure. In line with the literature, we find that the VRP, in absolute value, is more negative for long-term investments in low volatility periods. However, during high volatility periods, we uncover a nonmonotonic term structure of VRP, which reaches its maximum around a three-month maturity. This finding complements the recent results of Gruber et al. (2015) and Dew-Becker et al. (2017). 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 when conditioning on market conditions. It generates a Sharpe ratio of 0.77, compared to a meager ratio of 0.01 for a strategy that 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. We find that adding a term structure component substantially improves the predictive power of the VRP and explains our results in terms of spanning of the latent factors that describe variance.

Appendix A. Affine dependence of the VIX^2 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

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

$$\begin{aligned}\beta_{\mathrm{VIX}^2} = {} & \left(1 + 2\lambda_1^{(+)}C^{(+)} + 2\lambda_1^{(-)}C^{(-)}\right)B(\tau_{\mathrm{VIX}}) \\ & + \left(2\lambda_2^{(+)}C^{(+)} + 2\lambda_2^{(-)}C^{(-)}\right)\hat{A}(\tau_{\mathrm{VIX}}),\end{aligned} \tag{A.2}$$

$$\begin{aligned}\gamma_{\mathrm{VIX}^2} = {} & \left(1 + 2\lambda_1^{(+)}C^{(+)} + 2\lambda_1^{(-)}C^{(-)}\right)D(\tau_{\mathrm{VIX}}) \\ & + \left(2\lambda_3^{(+)}C^{(+)} + 2\lambda_3^{(-)}C^{(-)}\right)\bar{A}(\tau_{\mathrm{VIX}}),\end{aligned} \tag{A.3}$$

$$\begin{aligned}\delta_{\mathrm{VIX}^2} = {} & 2(\lambda_0^{(+)}C^{(+)} + \lambda_0^{(-)}C^{(-)}) \\ & + \left(1 + 2\lambda_1^{(+)}C^{(+)} + 2\lambda_1^{(-)}C^{(-)}\right)G +\end{aligned} \tag{A.4}$$

$$\begin{aligned}& \left(2\lambda_2^{(+)}C^{(+)} + 2\lambda_2^{(-)}C^{(-)}\right)\hat{B}(\tau_{\mathrm{VIX}}) \\ & + \left(2\lambda_3^{(+)}C^{(+)} + 2\lambda_3^{(-)}C^{(-)}\right)\bar{B}(\tau_{\mathrm{VIX}}),\end{aligned} \tag{A.5}$$

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

$$\begin{aligned}A(\tau) &= \frac{1}{\tau_{\mathrm{VIX}}}\int_t^{t+\tau_{\mathrm{VIX}}} \alpha_v(t,s)ds, \\ B(\tau) &= \frac{1}{\tau_{\mathrm{VIX}}}\int_t^{t+\tau_{\mathrm{VIX}}} \beta_v(t,s)ds,\end{aligned} \tag{A.6}$$

$$\begin{aligned}D(\tau) &= \frac{1}{\tau_{\mathrm{VIX}}}\int_t^{t+\tau_{\mathrm{VIX}}} \gamma_v(t,s)ds, \\ G(\tau) &= \frac{1}{\tau_{\mathrm{VIX}}}\int_t^{t+\tau_{\mathrm{VIX}}} \delta_v(t,s)ds,\end{aligned} \tag{A.7}$$

$$\begin{aligned}\hat{A}(\tau) &= \frac{1}{\tau_{\mathrm{VIX}}}\int_t^{t+\tau_{\mathrm{VIX}}} \alpha_m(t,s)ds, \\ \hat{B}(\tau) &= \frac{1}{\tau_{\mathrm{VIX}}}\int_t^{t+\tau_{\mathrm{VIX}}} \beta_m(t,s)ds,\end{aligned} \tag{A.8}$$

$$\begin{aligned}\bar{A}(\tau) &= \frac{1}{\tau_{\mathrm{VIX}}}\int_t^{t+\tau_{\mathrm{VIX}}} \alpha_u(t,s)ds, \\ \bar{B}(\tau) &= \frac{1}{\tau_{\mathrm{VIX}}}\int_t^{t+\tau_{\mathrm{VIX}}} \beta_u(t,s)ds.\end{aligned} \tag{A.9}$$

The functions α_v, β_v, γ_v, δ_v, α_m, β_m, α_u, and β_u are also available in closed form from:

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

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

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

for $0 \le t < s$.

Appendix B. Variance risk premium

The expression of the VRP involves the following terms:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left(\int_t^T v_s ds\right) = {} & A(T-t)v_t + B(T-t)m_t \\ & + D(T-t)u_t + G(T-t),\end{aligned} \tag{B.1}$$

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

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

where are A, B, D, G, $\hat{A}$, $\hat{B}$, $\bar{A}$, and $\bar{B}$ are given by Eqs. (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:

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}\left(\sum_{i=N_t^{(+/-)}}^{N_T^{(+/-)}} \left(Z_i^{(+/-)}\right)^2\right) &= \mathbb{E}_t^{\mathbb{Q}}\left(\sum_{i=N_t^{(+/-)}}^{N_T^{(+/-)}} 2\left(\mu_Y^{(+/-)}\right)^2\right) \\ &= 2\left(\mu_Y^{(+/-)}\right)^2\left(\lambda_0^{(+/-)}(T-t) + \lambda_1^{(+/-)}\mathbb{E}^{\mathbb{Q}}\left[\int_t^T v_s ds\right]\right. \\ &\quad \left. + \lambda_2^{(+/-)}\mathbb{E}^{\mathbb{Q}}\left[\int_t^T m_s ds\right] + \lambda_3^{(+/-)}\mathbb{E}^{\mathbb{Q}}\left[\int_t^T u_s ds\right]\right).\end{aligned} \tag{B.4}$$

Appendix C. Characteristic functions

The characteristic functions of the processes Y, VIX^2 are exponential affine in the state processes:

$$\Psi_{VIX_T^2}(t, v, m, u; \omega) = \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega VIX_T^2}\right] = e^{\alpha(T-t)+\beta(T-t)v+\gamma(T-t)m+\delta(T-t)u},$$

$$\Psi_{Y_T}(t, v, m, u; \omega) = \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega Y_T}\right] = e^{\alpha_Y(T-t)+\beta_Y(T-t)y+\gamma_Y(T-t)v+\delta_Y(T-t)m+\xi_Y(T-t)u},$$

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

$$-\alpha'(\tau) + \gamma(\tau)\kappa_m\theta_m + \delta(\tau)\kappa_u\theta_u + \lambda_0^{(-)}\left(\frac{1}{1-\beta(\tau)\nu_v^{(+)}} - 1\right) + \lambda_0^{(+)}\left(\frac{1}{1-\beta(\tau)\nu_v^{(-)}} - 1\right) = 0,$$

$$-\beta'(\tau) - \beta(\tau)\kappa_v + \frac{1}{2}\sigma_v^2\beta^2(\tau) + \lambda_1^{(-)}\left(\frac{1}{1-\beta(\tau)\nu_v^{(+)}} - 1\right) + \lambda_1^{(+)}\left(\frac{1}{1-\beta(\tau)\nu_v^{(-)}} - 1\right) = 0,$$

$$-\gamma'(\tau) - \gamma(\tau)\kappa_m + \frac{1}{2}\sigma_m^2\gamma^2(\tau) + \kappa_v\beta(\tau) + \lambda_2^{(-)}\left(\frac{1}{1-\beta(\tau)\nu_v^{(+)}} - 1\right) + \lambda_2^{(+)}\left(\frac{1}{1-\beta(\tau)\nu_v^{(-)}} - 1\right) = 0,$$

$$-\delta'(\tau) - \delta(\tau)\kappa_u + \frac{1}{2}\delta^2(\tau)\sigma_u^2 + \lambda_3^{(-)}\left(\frac{1}{1-\beta(\tau)\nu_v^{(+)}} - 1\right) + \lambda_3^{(+)}\left(\frac{1}{1-\beta(\tau)\nu_v^{(-)}} - 1\right) = 0,$$

$\forall t \in (0, T]$, with boundary conditions $\alpha(0) = \omega\delta_{VIX^2}$, $\beta(0) = \omega\alpha_{VIX^2}$, $\gamma(0) = \omega\beta_{VIX^2}$, 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.

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

$$-\alpha_Y'(\tau) + \beta_Y(\tau)\left(-\lambda_0^{(+)}\left(\frac{1}{1-\mu_Y^{(+)}} - 1\right) - \lambda_0^{(-)}\left(\frac{1}{1-\mu_Y^{(-)}} - 1\right)\right) + \delta_Y(T-t)\kappa_m\theta_m + \xi_Y(\tau)\kappa_u\theta_u$$
$$+ \lambda_0^{(+)}\left(\frac{1}{\left(1-\beta_Y(\tau)\mu_Y^{(+)}\right)\left(1-\gamma_Y(\tau)\nu_v^{(-)}\right)} - 1\right) + \lambda_0^{(-)}\left(\frac{1}{\left(1-\beta_Y(\tau)\mu_Y^{(-)}\right)\left(1-\gamma_Y(\tau)\nu_v^{(+)}\right)} - 1\right) = 0,$$

$$-\beta_Y'(\tau) = 0,$$

$$-\gamma_Y'(\tau) + \beta_Y(\tau)\left(-\lambda_1^{(+)}\left(\frac{1}{1-\mu_Y^{(+)}} - 1\right) - \lambda_1^{(-)}\left(\frac{1}{1-\mu_Y^{(-)}} - 1\right) - \frac{1}{2}\right) - \gamma_Y(\tau)\kappa_v + \frac{1}{2}\beta_Y(\tau)^2$$
$$+ \frac{1}{2}\gamma_Y(\tau)^2\sigma_v^2 + \beta_Y(\tau)\gamma_Y(\tau)\sigma_v\rho_{Y,v} + \lambda_1^{(+)}\left(\frac{1}{\left(1-\beta_Y(\tau)\mu_Y^{(+)}\right)\left(1-\gamma_Y(\tau)\nu_v^{(-)}\right)} - 1\right)$$
$$+ \lambda_1^{(-)}\left(\frac{1}{\left(1-\beta_Y(\tau)\mu_Y^{(-)}\right)\left(1-\gamma_Y(\tau)\nu_v^{(+)}\right)} - 1\right) = 0,$$

$$-\delta_Y'(\tau) + \beta_Y(\tau)\left(-\lambda_2^{(+)}\left(\frac{1}{1-\mu_Y^{(+)}} - 1\right) - \lambda_2^{(-)}\left(\frac{1}{1-\mu_Y^{(-)}} - 1\right)\right) + \gamma_Y(\tau)\kappa_v - \delta_Y(\tau)\kappa_m + \frac{1}{2}\delta_Y(\tau)^2\sigma_m^2$$
$$+ \lambda_2^{(+)}\left(\frac{1}{\left(1-\beta_Y(\tau)\mu_Y^{(+)}\right)\left(1-\gamma_Y(\tau)\nu_v^{(-)}\right)} - 1\right) + \lambda_2^{(-)}\left(\frac{1}{\left(1-\beta_Y(\tau)\mu_Y^{(-)}\right)\left(1-\gamma_Y(\tau)\nu_v^{(+)}\right)} - 1\right) = 0,$$

$$-\xi_Y'(\tau) + \beta_Y(\tau)\left(-\lambda_3^{(+)}\left(\frac{1}{1-\mu_Y^{(+)}} - 1\right) - \lambda_3^{(-)}\left(\frac{1}{1-\mu_Y^{(-)}} - 1\right)\right) - \xi_Y(\tau)\kappa_u + \frac{1}{2}\xi_Y(T-t)^2\sigma_u^2$$
$$+ \lambda_3^{(+)}\left(\frac{1}{\left(1-\beta_Y(\tau)\mu_Y^{(+)}\right)\left(1-\gamma_Y(\tau)\nu_v^{(-)}\right)} - 1\right) + \lambda_3^{(-)}\left(\frac{1}{\left(1-\beta_Y(\tau)\mu_Y^{(-)}\right)\left(1-\gamma_Y(\tau)\nu_v^{(+)}\right)} - 1\right) = 0,$$

with boundary conditions $\alpha_Y(0) = 0$, $\beta_Y(0) = \omega$, $\gamma_Y(0) = 0$, $\delta_Y(0) = 0$, and $\xi_Y(0) = 0$. The ODEs can be solved numerically using standard methods.

Appendix D. Coefficients for the Fourier cosine expansion

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

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

where $erfz(\cdot)$ is the error function for complex numbers. For $n = 0$,

$$
U_0^{\mathrm{VIX}^2} = \frac{2}{b-a}\left[\frac{2}{3}b^{3/2} - K_{\mathrm{VIX}}b + \frac{1}{3}K_{\mathrm{VIX}}^3\right]. \qquad \text{(D.2)}
$$

References

Aït-Sahalia, Y., Karaman, M., Mancini, L., 2012. The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis. Technical Report. Princeton University and Swiss Finance Institute.
Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. J. Financ. Econ. 83 (2), 413–452.
Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. J. Financ. 53 (2), 499–547.
Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. J. Financ. 57 (3), 1047–1091.
Amengual, D., 2008. The Term Structure Of Variance Risk Premia. Technical Report. Princeton University.
Amengual, D., Xiu, D., 2015. Resolution of Policy Uncertainty and Sudden Declines in Volatility. Technical Report. Centro de Estudios Monetarios y Financieros and University of Chicago.
Andersen, T., Fusari, N., Todorov, V., 2015. The risk premia embedded in index options. J. Financ. Econ. 117 (3), 558–584.
Andersen, T.G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. J. Financ. 57 (3), 1239–1284.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. J. Financ. Econ. 61, 43–76.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.
Andries, M., Eisenbach, T., Schmalz, M., Wang, Y., 2015. The Term Structure of the Price of Variance Risk. Technical Report. Federal Reserve Bank of New York.
Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. J. Financ. 52, 2003–2049.
Bali, T.G., Zhou, H., 2016. Risk, uncertainty, and expected returns. J. Financ. Quant. Anal. 51 (3), 707–735.
Bates, D.S., 1996. Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options. Rev. Financ. Stud. 9 (1), 69–107.
Bates, D.S., 2000. Post 87 crash fears in the S&P 500 futures option market. J. Econ. 94 (1–2), 181–238.
Bates, D.S., 2006. Maximum likelihood estimation of latent affine processes. Rev. Financ. Stud. 19 (3), 909–965.
Bates, D.S., 2012. US stock market crash risk, 1926–2010. J. Financ. Econ. 105 (2), 229–259.
Bayer, C., Gatheral, J., Karlsmark, M., 2013. Fast Ninomiya–Victoir calibration of the double-mean-reverting model. Quant. Financ. 13 (11), 1813–1829.
Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium, and stock market volatility. J. Econ. 183, 181–192.
Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Rev. Financ. Stud. 22, 4463–4492.
Bollerslev, T., Todorov, V., 2011. Tails, fears, and risk premia. J. Financ. 66 (6), 2165–2211.
Branger, N., Kraftschik, A., Völkert, C., 2016. The Fine Structure of Variance: Pricing VIX Derivatives in Consistent and Log-VIX Models. Technical Report. University of Muenster.
Broadie, M., Chernov, M., Johannes, M., 2007. Model specification and risk premia: evidence from futures options. J. Financ. 62 (3), 1453–1490.
Broadie, M., Chernov, M., Johannes, M., 2009. Understanding index option returns. Rev. Financ. Stud. 22 (11), 4493–4529.
Carr, P., Madan, D.B., 1999. Option valuation using the fast Fourier transform. J. Comput. Financ. 2 (4), 1–18.
Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G.T., 2003. Alternative models of stock price dynamics. J. Econ. 116 (1–2), 225–257.
Christoffersen, P., Heston, S., Jacobs, K., 2009. The shape and term structure of the index option smirk: why multifactor stochastic volatility models work so well. Manag. Sci. 55 (12), 1914–1932.
Christoffersen, P., Jacobs, K., Mimouni, K., 2010. Models for S&P 500 dynamics: evidence from realized volatility, daily returns, and options prices. Rev. Financ. Stud. 23 (8), 3141–3189.
Chung, S.-L., Tsai, W.-C., Wang, Y.-H., Weng, P.-S., 2011. The information content of the S&P 500 index and VIX options on the dynamics of the S&P 500 index. J. Futures Markets 31 (12), 1170–1201.
Cont, R., Kokholm, T., 2013. A consistent pricing model for index options and volatility derivatives. Math. Financ. 23 (2), 248–274.
Dew-Becker, I., Giglio, S., Le, A., Rodriguez, M., 2017. The price of variance risk. J. Financ. Econ. 123 (2), 225–250.
Drechsler, I., Yaron, A., 2011. What's vol got to do with it. Rev. Financ. Stud. 24, 1–45.
Duan, J.-C., Yeh, C.-Y., 2010. Jump and volatility risk premiums implied by VIX. J. Econ. Dyn. Control 34 (11), 2232–2244.
Duan, J.-C., Yeh, C.-Y., 2011. Price and Volatility Dynamics Implied by the VIX Term Structure. Technical Report. National University of Singapore.
Duffie, D., Pan, J., Singleton, K.J., 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. J. Financ. Econ. 11 (4), 650–681.
Egloff, D., Leippold, M., Wu, L., 2010. Valuation and optimal investing in variance swaps. J. Financ. Quant. Anal. 45 (5), 1279–1310.
Eraker, B., 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. J. Financ. 59 (3), 1367–1404.
Eraker, B., Johannes, M.S., Polson, N., 2003. The impact of jumps in equity index volatility and returns. J. Financ. 58, 1269–1300.
Fang, F., Oosterlee, C.W., 2008. A novel pricing method for European options based on Fourier–Cosine series expansions. SIAM J. Sci. Comput. 31 (2), 826–848.
Ferriani, F., Pastorello, S., 2012. Estimating and testing nonaffine option pricing models with a large unbalanced panel of options. Econ. J. 15 (2), 171–203. doi:10.1111/j.1368-423X.2012.00372.x.
Filipović, D., Gourier, E., Mancini, L., 2016. Quadratic variance swap models. J. Financ. Econ. 119 (1), 44–68.
Gatheral, J., 2008. Consistent modeling of SPX and VIX options. In: Proceedings of the Thirty-Seventh Bachelier Congress.
Gruber, P.H., Tebaldi, C., Trojani, F., 2015. The Price of the Smile and Variance Risk Premia. Technical Report. Swiss Finance Institute.
Hu, F., Zidek, J.V., 2002. The weighted likelihood. Can. J. Stat. 30 (3), 347–371.
Huang, D., Shaliastovich, I., 2015. Volatility-of-Volatility Risk. Technical Report. University of Pennsylvania.
Jacod, J., Todorov, V., 2010. Do price and volatility jump together? Ann. Appl. Prob. 20 (4), 1425–1469.
Jiang, G.J., Tian, Y.S., 2007. Extracting model-free volatility from option prices: an examination of the VIX index. J. Deriv. 14 (3), 35–60.
Johannes, M.S., Polson, N.G., Stroud, J.R., 2009. Optimal filtering of jump diffusions: extracting latent states from asset prices. Rev. Financ. Stud. 22 (7), 2759–2799.
Johnson, T.L., 2012. Equity Risk Premia and the VIX Term Structure. Technical Report. Stanford University.
Jones, C.S., 2003. The dynamics of stochastic volatility: evidence from underlying and options markets. J. Econ. 116 (1–2), 181–224.
Kaeck, A., Alexander, C., 2012. Volatility dynamics for the S&P 500: further evidence from non-affine, multi-factor jump diffusions. J. Bank. Financ. 36 (11), 3110–3121.
Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Rev. Financ. Stud. 27, 2841–2871.
Lindström, E., Ströjby, J., Brodén, M., Wiktorsson, M., Holst, J., 2008. Sequential calibration of options. Comput. Stat. Data Anal. 52 (6), 2877–2891.
Mencía, J., Sentana, E., 2013. Valuation of VIX derivatives. J. Financ. Econ. 108 (2), 367–391.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. J. Financ. Econ. 63 (1), 3–50.
Papanicolaou, A., Sircar, R., 2014. A regime-switching Heston model for VIX and S&P 500 implied volatilities. Quant. Financ. 14 (10), 1811–1827.
Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. J. Am. Stat. Assoc. 94 (446), 590–599.
Sepp, A., 2008. Pricing options on realized variance in the Heston model with jumps in returns and volatility. J. Comput. Financ. 11 (4), 33–70.
Song, Z., Xiu, D., 2016. A tale of two option markets: pricing kernels and volatility risk. J. Econ. 190 (1), 176–196.

Todorov, V., 2010. Variance risk premium dynamics: the role of jumps. Rev. Financ. Stud. 23 (1), 345–383.
Todorov, V., Tauchen, G., 2011. Volatility jumps. J. Bus. Econ. Stat. 29 (3), 356–371.
van Binsbergen, J.H., Hueskes, W., Koijen, R.S.J., Vrugt, E.B., 2013. Equity yields. J. Financ. Econ. 110 (3), 503–519.
Wu, L., 2011. Variance dynamics: Joint evidence from options and high--frequency returns. J. Econ. 160 (1), 280–287.