Asset Returns with Self-Exciting Jumps: Option Pricing and Estimation with a Continuum of Moments*

H. Peter Boswijk†
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

Roger J. A. Laeven‡
Amsterdam School of Economics
University of Amsterdam, EURANDOM
and CentER

Andrei Lalu§
Amsterdam School of Economics
University of Amsterdam, Tinbergen Institute
and Duisenberg School of Finance

February 5, 2016

Abstract

We propose an option pricing model with a self-exciting jump component inducing jump clustering. We develop a procedure to imply the latent state variables of the model from a panel of option prices and estimate the model parameters via the generalized method of moments employing a continuum of moments. Monte Carlo simulations show that our estimation procedure has good finite sample properties. Based on a panel of S&P 500 index option prices, we find strong evidence of self-excitation. The model's performance is shown to be superior to that of two alternative models with stochastic volatility and jumps.

Keywords: Self-exciting processes; option pricing; continuum of moments; conditional characteristic function; affine jump-diffusions.

JEL Classification: G12; C32; C58.

*_Matlab_ code to implement the estimation procedure developed in this paper is available from the authors upon request.

†Corresponding author. University of Amsterdam and Tinbergen Institute, P.O. Box 15867, 1001 NJ Amsterdam, The Netherlands. Email: H.P.Boswijk@uva.nl, Phone: +31 (0)20 5254252.

‡University of Amsterdam, EURANDOM and CentER, P.O. Box 15867, 1001 NJ Amsterdam, The Netherlands. Email: R.J.A.Laeven@uva.nl, Phone: +31 (0)20 5254219.

§University of Amsterdam, Tinbergen Institute and Duisenberg School of Finance, P.O. Box 15867, 1001 NJ Amsterdam, The Netherlands. Email: A.Lalu@uva.nl, Phone: +31 (0)20 5254245.

1 Introduction

A rich literature in financial econometrics documents the necessity of using both stochastic volatility and jumps to model asset returns and option price surfaces in continuous time. Yet when financial markets are in turmoil, asset price crashes occur more frequently than predicted by models with *standard* stochastic volatility and jump components, and the model-implied option price surface dynamics of standard models differ substantially from their empirical counterparts during such episodes. In periods of financial turmoil, asset and option price crashes tend to cluster over short time spans of days or even hours, and standard models are unable to capture this pattern of crash clustering. Therefore, alternative model specifications are required to better accommodate the empirical patterns found in asset returns, which leads to an improved fit for option pricing models.

Recently, Aït-Sahalia et al. (2015) introduced a novel way of modeling the complex interplay between crashes over time and across different equity markets by introducing mutually exciting jump processes, also known as Hawkes processes after Hawkes (1971), in a semi-martingale asset return model with a drift and stochastic volatility component. Aït-Sahalia et al. (2015) find evidence in favor of such a specification, with the self- and cross-excitation features that constitute mutual excitation in jumps being statistically significant in a panel data-set of global equity indices. Errais et al. (2010) employ a multivariate specification of the Hawkes point process to model clustered portfolio defaults. Hawkes processes are also used for the (high-frequency) modeling of trade book order arrivals and microstructure noise (e.g., Bowsher (2007) and Bacry et al. (2013)). Aït-Sahalia et al. (2014) use Hawkes processes to model contagion in default intensities implied from Eurozone sovereign CDS quotes.

We add to this literature by proposing an option pricing model with a self-exciting jump component inducing jump clustering, and by developing an estimation and testing procedure for our model that is novel in the setting with self-exciting processes. Specifically, the paper's contribution is fourfold. First, we extend extant option pricing models with Poissonian (hence Lévy) jumps by allowing for a self-exciting (hence non-Lévy) jump component, next to a standard stochastic volatility component. Second, we design a parametric estimation pro-

cedure for this model, which first involves backing out the model's latent states — stochastic volatility and jump intensity — from a panel of option prices and next employing a *continuum* of moments for estimation and testing, and we prove that, under natural assumptions, the resulting estimators are consistent. While our estimation procedure is developed in the context of self-exciting processes it can be applied more generally to jump-diffusion models satisfying our assumptions. Third, we apply our estimation method to equity index options, analyze the fit to the option price surface achieved by our model with self-exciting jumps, and illustrate the significant improvement compared to the fits achieved by several popular alternative jump-diffusion models. Finally, we analyze the risk premium embedded in our model, decompose the total risk premium in a diffusive risk premium and a jump risk premium, and design a fear gauge based on the difference between integrated quadratic variation under the physical probability measure (embedding risk premiums) and the risk neutral probability measure.

A main challenge to designing an identification strategy for our option pricing model with self-exciting jumps given a panel of option prices, comes from the fact that not only the stochastic volatility process but also the jump intensity process driving the jump component is inherently latent. To deal with this problem, we use option prices to first back out the parameter-dependent instantaneous volatility and jump intensity, i.e., the latent state variables, in a way that naturally generalizes the way in which volatility is implied from option prices in the standard Black and Scholes (1973) set-up. The implied jump intensity and volatility time series can then be used together with the asset return time series to estimate our model parameters. For a different option pricing model without self-excitation in jumps, a similar procedure of backing out states from option prices and next estimating the model parameters using the generalized method of moments (GMM) was employed by Pan (2002), who named this procedure implied-state GMM. It was further discussed by Pastorello et al. (2003) and Fan et al. (2015). Different from our approach, however, Pan (2002) uses standard GMM with a finite number of moments after having implied the latent states, rather than a continuum of moments as we do. This innovation will turn out to be not only theoretically but also practically highly relevant.

To build our continuum of moment conditions, we employ the conditional characteristic

function that we derive in closed form for our model. The use of conditional characteristic functions to estimate affine jump-diffusion models was proposed by Duffie et al. (2000) and Singleton (2001). A GMM estimator based on the conditional characteristic function is akin to the solution of an approximation to the first-order conditions of maximum likelihood. The number of points at which the conditional characteristic function is evaluated for the purpose of obtaining moment conditions can be extended from a finite number to a full continuum such that efficiency loss is minimized and the estimator becomes asymptotically efficient, as first shown by Carrasco and Florens (2002); see also Carrasco and Florens (2000) and Carrasco et al. (2007a). We adopt this approach and use a continuum of moment conditions derived from the conditional characteristic function to obtain our estimators. A key novelty that distinguishes our approach from earlier work on GMM estimation with a continuum of moments, is that we use option price inversion to imply latent states and accommodate the presence of latent variables: stochastic volatility and stochastic jump intensity.[1] We show that, under natural assumptions, our estimation procedure is consistent. We also show in extensive Monte Carlo simulations that our estimation procedure yields good performance in finite samples, even for the subtle characteristics of the jumps we are after.[2]

Due to the presence of a jump component with random jump size, our option pricing model induces an incomplete market with respect to the universe composed of the underlying stock, the finite number of option contracts and the money market account, meaning that the state price density is not unique. Choosing a suitable candidate pricing kernel that prices diffusive and jump risks,[3] the estimation procedure we develop provides a way to achieve the desirable identification and efficiency features of GMM estimation with a continuum of

[1] Carrasco et al. (2007a) suggest a possible adaptation of the simulated method of moments proposed by Duffie and Singleton (1993) to deal with the presence of latent state variables. Model parameters are then estimated from moment conditions based on the joint conditional characteristic function of observable state variables (e.g., forward returns) estimated via repeated simulations of (full) state vector paths. Our inversion approach avoids the need to resort to simulation.

[2] In notable recent work Andersen et al. (2015a) propose a joint estimation of parametric models and recovery of latent states variables in the context of option panels observed with errors. Their estimation procedure relies on a penalized non-linear least squares approach which uses additional information from high-frequency returns on the option underlying, requiring no parametric assumptions about the physical measure governing the state variables. Our approach does not use high-frequency data.

[3] Other risk factors such as interest rate uncertainty, dividend uncertainty or liquidity concerns could potentially have a separate impact on option prices. We focus on jump and diffusive risks to investigate the differences between and the dynamic behavior of the risk premiums originating from these two risk sources.

moment conditions while simultaneously estimating the model under both the physical and the risk neutral probability measures, and the risk premiums that link the two.

We implement our estimation procedure on a rich panel of S&P 500 options prices and find strong evidence of self-excitation in jumps. Furthermore, we show that the model is better capable of reproducing option price surfaces than alternative models with (possibly) time-varying jump intensities, both in-sample and out-of-sample. Not only does the model provide a good fit to option price surfaces within a trading day, but, thanks to the self-exciting feedback property, it can also fit different shapes and profiles of the price surface throughout different trading days and across different market regimes, using the same time-series-based estimates of the model parameters.

The data-set we use for our empirical application consists of prices of European option contracts on the S&P 500 index traded in the time period between January 1996 and August 2013. In our implementation we use, for each time point, a rich set of option prices, with both put and call contracts and multiple maturities and money-ness levels. Thus, the latent states are not implied from a randomly picked pair of options at each time point, but in a uniform way throughout the whole length of the data sample. Both the parameter estimates and the implied latent states obtained support the hypothesis of a stochastic jump intensity that allows the jumps in the S&P 500 index returns to self-excite, hence exhibiting a pattern of jump clustering.

Our model specification with self-exciting jumps constitutes an alternative parametric approach to explaining the observed differences between the time-series behavior of (risk neutral) asset return distributions implied from option prices and the time-series behavior of (physical) asset return distributions derived from actual prices of the underlying asset. Existing studies,[4] based on either parametric or non-parametric techniques, find statistically significant evidence of a jump risk premium component embedded in equity returns. Most of the existing work is geared towards establishing the existence and gauging the (relative) size of jump risk premiums. Some of these studies also employ state-dependent intensities for the jump process, however the intensity of the jump process is commonly linked to other state or exogenous variables (e.g.,

[4] Among the papers dedicated to this topic are e.g., Bates (2000), Pan (2002), Eraker (2004), and Bollerslev and Todorov (2011).

to the stochastic volatility process). By contrast, in our approach, the jump process and the associated jump risk premium are linked to their own history — an inherent feature of Hawkes point processes — leading to a realistic and parsimonious approach.

After a first jump occurs, the risk premium investors demand to bear the risk of future jumps occurring spikes up and afterwards, if no subsequent jumps occur, it decreases, with the average half-life of an asset price jump shock to the intensity process being approximately two weeks, according to our estimates. During calm market conditions the diffusive risk premium plays a greater role than the jump risk premium, but the latter immediately becomes dominant after a jump occurs. Based on our model's ability to capture time-varying jump risk patterns, we develop a parsimonious investor fear gauge derived from integrated quadratic variation. The proposed gauge spikes when extreme market movements associated with jumps occur and clearly delineates periods of market turmoil attributable to investor fears of (subsequent) large price shocks.

The remainder of the paper is structured as follows: Section 2 describes the properties of the Hawkes self-exciting jump process we employ, specifies the asset return dynamics and the candidate pricing kernel, introduces the option pricing model and derives the conditional characteristic function for our model in closed form. Section 3 describes our parametric estimation procedure, proves its consistency, and presents our Monte Carlo results. Section 4 contains our empirical analysis and assesses the model performance compared to alternative option pricing models. Section 5 analyzes model-based risk premiums and the investor fear gauge we propose. Section 6 concludes the paper. Proofs and some additional details, such as the descriptive statistics and the description of alternative option pricing models, are relegated to the Appendix. The paper is supplemented by a separate online appendix[5] containing further details on the Monte Carlo analysis, the identification of the latent states and of the model parameters, and the adopted numerical methods.

[5]Available from the authors' webpages; see `http://www.rogerlaeven.com`.

2 The Model

2.1 Self-Exciting Jump Processes

Throughout we consider a filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ satisfying the usual conditions.[6] To model the dynamics of the jump component in our option pricing model, we employ a Hawkes self-exciting point process (Hawkes (1971)), defined on this probability space. The Hawkes process, denoted by the pair (N, λ), where N is a counting process with stochastic jump intensity λ, is a path-dependent point process and can be restricted to become a Markov process (as we will do below). The stochastic jump intensity λ_t describes the $\mathcal{F}_t$-conditional instantaneous mean jump rate of N_t per unit of time, i.e.:

$$\begin{cases} \mathbb{P}\left[N_{t+\Delta} - N_t = 0 | \mathcal{F}_t\right] = 1 - \lambda_t \Delta + o(\Delta); \\ \mathbb{P}\left[N_{t+\Delta} - N_t = 1 | \mathcal{F}_t\right] = \lambda_t \Delta + o(\Delta); \\ \mathbb{P}\left[N_{t+\Delta} - N_t > 1 | \mathcal{F}_t\right] = o(\Delta); \end{cases} \tag{2.1}$$

with $\Delta > 0$ a small time step.

The characteristic feature of the Hawkes process is its self-exciting property, which induces an interdependence between the processes N and λ. This feature is best described by the stochastic integral equation which specifies the dynamics of the stochastic jump intensity process λ. Under exponential decay, it is given by

$$\lambda_t = \bar{\lambda} + \int_{-\infty}^{t} \delta e^{-\kappa_\lambda (t-s)} \, \mathrm{d}N_s. \tag{2.2}$$

Then, the pair (N, λ) is Markov. If we interpret (2.2) in the context of asset price jumps, $\bar{\lambda} > 0$ represents the intensity until the first (ever) asset price jump occurs, while $\delta e^{-\kappa_\lambda (t-s)}$, with $\kappa_\lambda > \delta \geq 0$, measures the (decaying) impact of an asset price jump onto the intensity process. Thus, the intensity driving N is given by λ and, in turn, λ also changes in response to an increase in the counting process N: jumps feed back into the intensity process which governs

[6] See e.g., Protter (2005), p. 3.

the likelihood of future jumps, which constitutes self-excitation. The restriction $\kappa_\lambda > \delta \geq 0$ ensures that the intensity process is positive with probability one and (conditionally) stationary. The impact of the occurrence of jumps decays over time at an exponential rate κ_λ to the steady state level $\bar{\lambda}$. The compensated process $N_t - \int_{-\infty}^{t} \lambda_s \, \mathrm{d}s$ is a local martingale.

One readily verifies that the intensity process (2.2) has the equivalent differential formulation:

$$\mathrm{d}\lambda_t = \kappa_\lambda(\bar{\lambda} - \lambda_t) \ \mathrm{d}t + \delta \, \mathrm{d}N_t. \tag{2.3}$$

Here the drift term $\kappa_\lambda(\bar{\lambda} - \lambda_t) \, \mathrm{d}t$ generates mean reversion (to $\bar{\lambda}$ at rate κ_λ) and the jump component $\delta \, \mathrm{d}N_t$ generates self-excitation. By the self-exciting property, the Hawkes point process exhibits positive autocorrelation. Thus, the Hawkes self-exciting point process falls outside the class of Lévy processes, which is characterized by the property of serially independent increments. To illustrate the typical dynamics of the Hawkes process, we display some simulated sample paths of the pair (N, λ) in Appendix A.

2.2 The Equity Return Dynamics under $\mathbb{P}$

Next, we introduce our model for equity return dynamics. It may be viewed as an extension of the classic Bates (2000) model. On our filtered probability space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\big)$, with $\mathbb{P}$ the physical (historical) probability measure, we assume that the log-forward price process $y_t = \log(F_t)$,[7,8] the stochastic volatility v_t, and the stochastic jump intensity λ_t of the point

[7] $F_t, t \geq 0$, denotes the forward price process of a standard forward contract traded on an index or a stock for a given fixed maturity. We assume that the absence of cash-and-carry arbitrage opportunities ensures that forward price processes for different forward contract maturities have the same dynamics (but different initial values).

[8] By modeling the log-forward returns instead of the log-returns we avoid having to deal explicitly with dividend yields (and, it further ensures in our empirical analysis that the log-forward observations we use are synchronized with the option price observations by backing out the log-forward observations from option prices using the put-call parity relationship).

process with counter N_t, have the following dynamics:

$$\mathrm{d}y_t = \left(\left(\eta - \frac{1}{2}\right) v_t + (\mu - \mu^*)\lambda_t\right) \mathrm{d}t + \sqrt{v_t}\, \mathrm{d}W_t^{(1),\mathbb{P}} + \mathrm{d}J_t^{\mathbb{P}} - \mu\lambda_t\, \mathrm{d}t; \tag{2.4}$$

$$\mathrm{d}v_t = \kappa_v\left(\overline{v} - v_t\right) \mathrm{d}t + \sigma_v\sqrt{v_t}\left(\rho\, \mathrm{d}W_t^{(1),\mathbb{P}} + \sqrt{1-\rho^2}\, \mathrm{d}W_t^{(2),\mathbb{P}}\right); \tag{2.5}$$

$$\mathrm{d}\lambda_t = \kappa_\lambda\left(\overline{\lambda} - \lambda_t\right) \mathrm{d}t + \delta\, \mathrm{d}N_t. \tag{2.6}$$

The model (2.4)–(2.6) is designed to capture two salient features of equity return dynamics: stochastic volatility and clustered price jumps. Stochastic volatility is modeled through the local variance process v_t of Heston (1993) defined in (2.5), which is also known as the Feller (1951) square-root process or the CIR process after Cox et al. (1985). Here $\left(W_t^{(1),\mathbb{P}}, W_t^{(2),\mathbb{P}}\right)$ are $\mathcal{F}_t$-adapted standard Brownian motions under the physical probability measure $\mathbb{P}$. The process v_t is mean reverting towards a long run mean $\overline{v} > 0$ at a mean reversion rate $\kappa_v > 0$. To ensure the process is strictly positive we impose the Feller condition $2\kappa_v\overline{v} > \sigma_v^2 > 0$. The specification further renders the Brownian component increments in (2.5) to be correlated with the Brownian component increments in (2.4), via the correlation parameter ρ, $-1 \leq \rho \leq 1$, capturing the leverage effect, coined by Black (1976).

Pertaining to the jump component in our model, $\mathrm{d}J_t^{\mathbb{P}}$ denotes a compound Hawkes jump increment. More specifically, we let $\mathrm{d}J_t^{\mathbb{P}} = Z_t^{\mathbb{P}}\, \mathrm{d}N_t$, with $Z_t^{\mathbb{P}}$ a serially independent random variable governing the jump size and $\mathrm{d}N_t$ the instantaneous increment of the Hawkes self-exciting jump counter with jump intensity given by (2.6). Conditionally upon the arrival of a jump, the forward price jump is $F_t - F_{t-} = F_{t-}\left(\exp(Z_t^{\mathbb{P}}) - 1\right)$. We assume that $Z_t^{\mathbb{P}}$ is normally distributed with mean $\mu_j^{\mathbb{P}}$ and standard deviation σ_j. This implies that the mean relative jump size of the forward price under the physical measure $\mathbb{P}$ is $\mu = \mathbb{E}\left[\exp(Z_t^{\mathbb{P}}) - 1\right] = \exp\left(\mu_j^{\mathbb{P}} + \frac{1}{2}\sigma_j^2\right) - 1$. Using a log-normal distribution to model the distribution of equity return jump sizes is a canonical choice in the literature concerned with the modeling of equity price jumps, dating back to the seminal paper by Merton (1976). The jump times in our model are determined by the counting process N_t with intensity λ_t, the pair forming the Hawkes process detailed in the previous section. We assume that the jump size Z and the counting process N are mutually independent. Furthermore, we assume that the vector of Brownian motions is

independent of the stochastic components in the jump process.

Some studies (e.g., Eraker et al. (2003) and Eraker (2004)) propose the inclusion of a jump component in the volatility process, possibly synchronizing the jumps in volatility with the ones in the equity price process. Other studies (e.g., Egloff et al. (2010)) argue in favor of using a two-factor stochastic volatility model with jumps, allowing the volatility process to revert towards a stochastic central tendency (modeled as a square-root process, possibly with jumps). In these papers, the main factor driving the dynamics of the option-implied volatility surface is the variance process. The motivation for these model specifications is in part due to the need to better fit volatility dynamics over time. We do not consider any of these variations in the local variance process (2.5), for two reasons. First, as we will see, the square-root local variance process together with the self-exciting jump process in returns is already quite flexible in accounting for the dynamics of the option price surface.[9] Second, although it would no doubt be appealing to accommodate all these effects, this will necessarily lead to econometric identification problems, given the number of latent processes involved and the subtle characteristics we are already after. Therefore, we restrict to a parsimonious specification in (2.5).

The drift component in (2.4) will be discussed in detail in the following subsection.

2.3 Candidate Pricing Kernel and Market Prices of Risk

Given the model (2.4)–(2.6), the financial market consisting of the underlying stock, a finite number of option contracts on the stock, and a money market account is rendered incomplete. As a result, the pricing kernel (or state price density) for arbitrage-free valuation is not uniquely defined. We restrict our attention to a candidate pricing kernel that keeps the dynamics of the log-forward, stochastic volatility and stochastic jump intensity processes under the pricing (risk neutral or equivalent martingale) measure $\mathbb{Q}$ within the affine jump-diffusion model class.[10] Inspired by Pan (2002), Eraker (2004) and Broadie et al. (2007), we consider a state price

[9] Andersen et al. (2015b) also document the importance of modeling jump tail dynamics separately from the volatility dynamics in order to capture salient features of the implied volatility surface dynamics.

[10] For a formal definition of the affine jump-diffusion model class, we refer to Duffie et al. (2000).

density, ξ_t, of the following form:

$$\xi_t = \exp\Big(- \int_0^t r_s \, \mathrm{d}s \Big) \xi_t^D \xi_t^J; \tag{2.7}$$

$$\xi_t^D = \exp\Big[- \Big(\int_0^t \Gamma_{1,s} \, \mathrm{d}W_s^{(1),\mathbb{P}} + \int_0^t \Gamma_{2,s} \, \mathrm{d}W_s^{(2),\mathbb{P}} \Big) - \frac{1}{2} \Big(\int_0^t \Gamma_{1,s}^2 \, \mathrm{d}s + \int_0^t \Gamma_{2,s}^2 \, \mathrm{d}s \Big) \Big]; \tag{2.8}$$

$$\xi_t^J = \exp\Big(\sum_{s_i^* \leq t} Z_{s_i^*}^* \Big); \tag{2.9}$$

with r the risk-free interest rate process induced by the money market account.

In ξ_t^D, which is the part of the state price density that governs the market prices of risk for the diffusive and volatility components, the processes Γ_1 and Γ_2 are defined as:

$$\Gamma_{1,t} = \eta \sqrt{v_t}, \qquad \Gamma_{2,t} = - \frac{\rho}{\sqrt{1-\rho^2}} \left(\eta + \frac{\eta^v}{\sigma_v} \right) \sqrt{v_t}. \tag{2.10}$$

The market price of diffusive risk in the log-forward process, $\Gamma_{1,t}$, is determined by the parameter η. It captures a trade-off akin to the risk-return trade-off in the CAPM framework: the larger η, the higher the premium for diffusive risk. This market price of risk is time-varying, because the volatility process, v_t, changes stochastically over time, hence the risk premium per unit of time, given by ηv_t, also changes over time. The second risk premium parameter, η^v, determines the market price of diffusive risk in the volatility process, known as the market price of volatility risk. There is no clear empirical evidence concerning the statistical significance, the magnitude, or the sign of the diffusive volatility risk premium parameter η^v in empirical studies on equity index returns using models with a stochastic volatility state.[11] We therefore opt to constrain the parameter η^v to be zero and do not account for a volatility risk premium in the volatility process.

The state price density pertaining to the jump component, ξ_t^J, is dictated by the random variables $Z_{s_i^*}^*$, which are assumed to be i.i.d. normally distributed with mean μ_j^* and standard deviation σ_j^*. The time points s_i^* denote the times at which the forward price experiences a

[11] See e.g., Broadie et al. (2007) for a discussion of the empirical evidence on the size and sign of the volatility risk premium.

jump. This means that ξ_t^J exhibits jumps simultaneously with the underlying asset, i.e., ξ_t^J takes effect when the underlying asset jumps as well. The $Z^*_{s_i^*}$ are assumed to be independent from all other stochastic components in the model, such as the Brownian motions, the jump times, and all previous jump sizes in the kernel and in the log-forward price, except for the jump size variable $Z_{s_i^*}$ in the log-forward price, with which they are correlated contemporaneously. The stochastic behavior of the $Z^*_{s_i^*}$ governed by their mean and variance together with their correlation with $Z_{s_i^*}$ determine the jump-timing and jump-size risk premiums of the model. We impose the restriction that $\mu_j^* + \frac{1}{2}(\sigma_j^*)^2 = 0$, which constrains the jump risk premiums to be jump-size related and not jump-timing related; see Appendix B for further details.

As in Pan (2002), only the mean of the jump size distribution changes between the risk neutral specification and the physical probability specification of the model, i.e., $\mu_j^{\mathbb{P}} \neq \mu_j^{\mathbb{Q}}$. The jump risk premium originates from this difference between the means of the jump size distributions in the log-forward returns. As a consequence, the mean relative jump size under $\mathbb{P}$, μ, also differs from the mean relative jump size under $\mathbb{Q}$, μ^*. The candidate pricing kernel restricts the jump size variance σ_j^2 to be the same under both measures. Given this specification of the state price density, the expected excess return (risk premium) that investors demand in exchange for bearing asset price jump risks is $(\mu - \mu^*)\lambda_t$ per unit of time. This premium is time-varying, because the latent jump intensity process λ_t changes over time.

These choices related to the diffusive and jump part of the state price density are imposed to keep model estimation tractable under both probability measures. A richer specification could be made, but this would come at the expense of weaker parameter identification. In part, the identification problem stems from the classical peso problem: as jumps are rare events, pinning down estimates for their distributions (under both probability measures) is challenging.

Finally, we note that if we define the density process $\xi_t \exp\Big(\int_0^t r_s \,\mathrm{d}s\Big)$, we obtain a local martingale that can be used to generate an equivalent martingale measure $\mathbb{Q}$. Under this probability measure $\mathbb{Q}$, we let $W_t^{(i),\mathbb{Q}} = W_t^{(i),\mathbb{P}} + \int_0^t \Gamma_{i,s} \,\mathrm{d}s$ for $i = 1, 2$. Further details on the pricing kernel and the measure change it induces are provided in Appendix B.

2.4 The Equity Return Dynamics under $\mathbb{Q}$

Under the equivalent martingale measure $\mathbb{Q}$ associated with the chosen candidate pricing kernel ξ_t, the log-forward return process has the following dynamics:

$$\mathrm{d}y_t = -\frac{1}{2}v_t\,\mathrm{d}t + \sqrt{v_t}\,\mathrm{d}W_t^{(1),\mathbb{Q}} + \mathrm{d}J_t^{\mathbb{Q}} - \mu^*\lambda_t\,\mathrm{d}t; \tag{2.11}$$

$$\mathrm{d}v_t = \kappa_v\left(\overline{v} - v_t\right)\mathrm{d}t + \sigma_v\sqrt{v_t}\left(\rho\,\mathrm{d}W_t^{(1),\mathbb{Q}} + \sqrt{1-\rho^2}\,\mathrm{d}W_t^{(2),\mathbb{Q}}\right); \tag{2.12}$$

$$\mathrm{d}\lambda_t = \kappa_\lambda\left(\overline{\lambda} - \lambda_t\right)\mathrm{d}t + \delta\,\mathrm{d}N_t. \tag{2.13}$$

Here $(W_t^{(1),\mathbb{Q}}, W_t^{(2),\mathbb{Q}})$ are standard Brownian motions under $\mathbb{Q}$ and $\mathrm{d}J_t^{\mathbb{Q}} = Z_t^{\mathbb{Q}}\,\mathrm{d}N_t$. The jump intensity process is the same as under $\mathbb{P}$, i.e., the counting process N_t is not affected by the measure change. Conditionally upon a jump event taking place, the risk neutral mean relative jump size is $\mu^* = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(Z_t^{\mathbb{Q}}\right) - 1\right] = \exp\left(\mu_j^{\mathbb{Q}} + \frac{1}{2}\sigma_j^2\right) - 1$. Jumps under the risk neutral measure have a different mean than under the physical measure ($\mu_j^{\mathbb{P}} \neq \mu_j^{\mathbb{Q}}$). The last term in (2.11) is the corresponding jump compensator for the jump process under $\mathbb{Q}$, such that the log-forward return is a (local) martingale under $\mathbb{Q}$.

2.5 Closed-Form Option Pricing with Self-Excitation

In both its specification under $\mathbb{P}$ and under $\mathbb{Q}$, the model falls into the affine jump-diffusion class of models in its generalized version as defined in Appendix B of Duffie et al. (2000). For affine models, option pricing is efficiently done using Fourier inversion techniques, given that the conditional characteristic function of the vector of stochastic states can be derived. So, in our set-up, we derive the conditional characteristic function of $(y_T, v_T, \lambda_T)^\top$ — the state vector — and invert it to find the price of the derivative contract at the current time t, with $0 \leq t < T$, assuming (for now) that the current state vector $(y_t, v_t, \lambda_t)^\top$ and the model parameters are known. The computation of the conditional characteristic function can be carried out in closed form up to the solution of a system of ordinary differential equations (ODEs). Its derivation is based on the results obtained in the general setting of affine jump-diffusions developed in Duffie et al. (2000).

Specifically, the system of stochastic differential equations driving the state vector $X_t \coloneqq (y_t, v_t, \lambda_t)^\top$ of our model under $\mathbb{Q}$ can be written in the following matrix form:

$$\underbrace{\mathrm{d}\begin{pmatrix} y_t \\ v_t \\ \lambda_t \end{pmatrix}}_{\coloneqq \mathrm{d}X_t} = \underbrace{\begin{pmatrix} -\frac{v_t}{2} - \mu^* \lambda_t \\ \kappa_v(\overline{v} - v_t) \\ \kappa_\lambda(\bar{\lambda} - \lambda_t) \end{pmatrix} \mathrm{d}t}_{\coloneqq \mu(X_t)\,\mathrm{d}t}$$
$$+ \underbrace{\begin{pmatrix} \sqrt{v_t} & 0 & 0 \\ \sigma_v \rho \sqrt{v_t} & \sigma_v \sqrt{1-\rho^2}\sqrt{v_t} & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{\coloneqq \sigma(X_t)} \underbrace{\begin{pmatrix} \mathrm{d}W_t^{(1),\mathbb{Q}} \\ \mathrm{d}W_t^{(2),\mathbb{Q}} \\ 0 \end{pmatrix}}_{\coloneqq \mathrm{d}W_t^{3\times 1}} + \underbrace{\begin{pmatrix} Z_t^{\mathbb{Q}} \\ 0 \\ \delta \end{pmatrix} \mathrm{d}N_t}_{\coloneqq Z_t^{3\times 1}\,\mathrm{d}N_t}, \tag{2.14}$$

adopting the notation of Duffie et al. (2000). This system can be restricted to fit into the class of affine jump-diffusions (in its generalized version appearing in Appendix B of Duffie et al. (2000)), as $\mu(\cdot)$ and $\sigma(\cdot)\sigma(\cdot)^\top$ are affine functions of their argument X_t, and (2.14) can be written as:

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t^{3\times 1} + Z_t^{3\times 1}\,\mathrm{d}N_t. \tag{2.15}$$

Given the current state vector X_t at time $t \geq 0$ and the model parameter set which we denote by θ, the conditional characteristic function of the state vector at time $T > t$ under $\mathbb{Q}$ with argument $u \in \mathbb{C}^3$ is defined as:

$$\phi^{\mathbb{Q}}(u, X_t, T-t; \theta) = \mathbb{E}^{\mathbb{Q}}\left[e^{\mathrm{i}u \cdot X_T} | \mathcal{F}_t\right]. \tag{2.16}$$

Proposition 1 *The conditional characteristic function of the state vector under $\mathbb{Q}$ is given by:*

$$\phi^{\mathbb{Q}}(u, X_t, T-t; \theta) = e^{\alpha(T-t) + \beta_1(T-t)y_t + \beta_2(T-t)v_t + \beta_3(T-t)\lambda_t},$$

where $\alpha(T-t)$ *and* $\beta_i(T-t)$ *are solutions to the following system of ODEs:*

$$\begin{aligned}
\dot{\beta}_1 &= 0 \\
\dot{\beta}_2 &= -\frac{\beta_1}{2}(1-\beta_1) - \kappa_v\beta_2 + \sigma_v\rho\beta_1\beta_2 + \frac{1}{2}\sigma_v^2\beta_2^2 \\
\dot{\beta}_3 &= -\left(e^{\mu_j^{\mathbb{Q}}+\frac{\sigma_j^2}{2}} - 1\right)\beta_1 - \kappa_\lambda\beta_3 + e^{\beta_1\mu_j^{\mathbb{Q}}+\frac{\beta_1^2\sigma_j^2}{2}+\delta\beta_3} - 1 \\
\dot{\alpha} &= \kappa_v\bar{v}\beta_2 + \kappa_\lambda\bar{\lambda}\beta_3,
\end{aligned}$$

subject to the initial conditions $\beta(T-0) = (u_1, u_2, u_3)^\top$ *and* $\alpha(T-0) = 0$.

The proof of Proposition 1 follows from the application of Proposition 1 in Appendix B of Duffie et al. (2000). The system of ODEs appearing in the proposition above can be solved using available numerical methods (e.g., Runge-Kutta), however a fully analytic solution is not possible due to the non-linearities in the differential equation involving $\dot{\beta}_3$.[12]

With the conditional characteristic function for the log-forward price given in Proposition 1 above,[13] we use the Fast Fourier Transform approach of Carr and Madan (1999) to price European options. We denote call option contract maturity and (forward) money-ness by $[T,k] \in \mathbb{R}^2_+$, where $T \geq t$ and $k = K/F_t$, with K the strike price.[14] At any particular time point t we are interested in the prices of several call options with different maturities and money-ness levels, therefore we use the index $q = 1, \ldots, \overline{q},\ \overline{q} \in \mathbb{N}^+$ to collect all the characteristic pairs $[T,k]$ for these contracts in a matrix $C = \left(C_1, C_2, \ldots, C_q, \ldots, C_{\overline{q}}\right)^\top$.

Given the state vector X_t and the model parameters θ, we define the option pricing function that determines the price vector $p_t = \left(p_t(C_1), \ldots, p_t(C_q), \ldots, p_t(C_{\overline{q}})\right)^\top$ containing the time t prices for European call[15] options by:

$$p_t = P(X_t, \theta, C). \tag{2.17}$$

[12] The first equation is trivially solved by $\beta_1(T-t) = u_1$, while a fully analytic solution to the second one which falls into the class of Ricatti differential equations can be found in Duffie et al. (2000).

[13] That is, by evaluating the system of ODEs with the initial conditions $\beta(T-0) = (u_1, 0, 0)^\top$ and $\alpha(T-0) = 0$.

[14] F_t denotes the forward price for a given reference maturity.

[15] Put option prices can be obtained by using the standard put-call parity for European options.

3 The Estimation Procedure

In this section we describe our strategy to estimate the model parameter set θ. The data-set that our estimation procedure employs is a rich panel of European option prices. A main challenge when devising an estimation procedure for our model is provided by the fact that two of the three variables in the state vector X are latent: we only have (discretely sampled) observations of the log-forward returns in X,[16] while the stochastic volatility and jump intensity processes are not observable.

One possible approach to estimate the model parameters is to minimize price differences between the market-observed option prices and the model-implied option prices, by varying the parameters and the levels of the stochastic latent states. This approach, sometimes referred to as "calibration", is often adopted for option pricing with stochastic volatility models, in applications where the main research objective is to fit a model to a snapshot of the daily market-implied volatility surface. In our case, such an approach would yield point values for the parameters of the model under the risk neutral specification, but would not give any insights about the risk premiums linking the two probability measures nor about the longer horizon dynamics of the return distribution under the physical probability measure. By exploiting the time-series dimension of our panel of option prices,[17] we can jointly estimate the $\mathbb{P}$ and $\mathbb{Q}$ dynamics of our model. We now briefly outline the main intuition behind our estimation procedure; the full details about each estimation step are presented in the next subsections.

If we could observe the full state vector X at a set of time points $t_1, t_2, \ldots, t_i, \ldots, t_n, t_{n+1}$, we could use the equations for the model dynamics under $\mathbb{P}$ to write down a set of $\mathcal{F}_{t_i}$-conditional moment conditions that would contain and identify all model parameters, including the risk premium parameters η and $\mu_j^{\mathbb{P}}$, which appear in the model's $\mathbb{P}$-dynamics. However, we do not observe the full state vector. In order to deal with the latency of the processes v_t and λ_t we use market quoted option prices to infer the levels of the latent state processes at the dis-

[16] To avoid non-stationarity problems in the series that we base our estimation upon, we substitute the log-forward price levels by log-forward returns in the state vector X_t. We do not explicitly account for this in our notation, but this substitution applies throughout this entire section.

[17] We use near at-the-money put and call options to determine the forward price which can serve as an underlying for an option contract written on the forward price. The time series of log-forward returns is used in lieu of the log-index returns for practical considerations which are detailed in the next section.

crete time points at which the option prices are quoted, and conduct estimation conditionally upon the implied states.

This type of estimation approach by first backing out latent states from option prices and next estimating model parameters was also used in Pan (2002). In her paper, the jump intensity is allowed to time-vary, but is restricted to be a multiple of the volatility process, i.e., $\lambda_t = \lambda_1 v_t,\ \lambda_1 \in \mathbb{R}^+$, which implies that the state vector contains a single latent variable, the unobservable stochastic volatility. Compared to our set-up, the model in Pan (2002) imposes a much stronger assumption on the jump intensity process, whose stochastic component is a deterministic multiple of the stochastic component generated by the two Brownian components in the volatility process. In particular, there is no feedback between jumps and the stochastic jump intensity, while this self-exciting feature is among the key interests in this paper. We will compare the performance of the two models in detail in Section 4. Furthermore, different from our approach, Pan (2002) applies standard GMM with a finite number of moment conditions after having implied the latent states. We will use a *continuum* of moments to conduct estimation. Using a continuum of moments in implied-state GMM is not just a theoretical innovation compared to Pan (2002), but it is also practically relevant as the approach of Pan (2002) would yield computational and finite sample identification problems for our model; see Section 3.3 for further details.

Following Black-Scholes, the concept of implied volatility refers to the backing out of the unobservable constant volatility parameter from market prices of traded derivative contracts. The concept has been extended to models with a stochastic volatility latent factor and can further be extended to a two latent factors model set-up, such as the one we are considering. As opposed to the Black-Scholes case in which the implied volatility does not depend on any other parameters from the Black-Scholes model, in our case, the model-implied state variables depend on the parameter vector θ. When implied from option prices at the true parameter vector, say θ_0, the implied states will be different from when they are implied from a different parameter vector, say θ_1. If the model is correctly specified, then, given that the states implied from a parameter vector different from the true one do not follow the dynamics imposed by that parameter vector (we show that this happens only for θ_0), we can devise an estimation

strategy for the model parameters, which happens intuitively by ruling out implied state series that are unlikely under our assumed dynamics.

Starting from an initial set of parameters θ, we use the option pricing framework of Section 2.5 to first back out v_t^θ and λ_t^θ. Next, we evaluate a continuum of moment conditions derived from the model under $\mathbb{P}$ based on the full state vector $\left(y_t, v_t^\theta, \lambda_t^\theta\right)^\top$ and the initial choice of θ. Then, we change θ and repeat the process again, iterating until a moment based criterion function is minimized. The parameter set that minimizes the moment based criterion function (for a given sample) is an estimator of the true parameter set θ_0. It will be proven to be a consistent estimator. We now describe each step in the estimation procedure in further detail.

3.1 Backing Out Latent States

Assume that market-quoted prices are observed with a regular frequency Δ at equally spaced[18] time points $\{t_1, t_2, \ldots, t_i, \ldots, t_n, t_{n+1}\}$ for the log-forward price y_{t_i} and for European-style derivative contracts $p_{t_i} = \left(p_{t_i}(C_1), \ldots, p_{t_i}(C_q), \ldots, p_{t_i}(C_{\overline{q}})\right)^\top$ with $\overline{q} \geq 2$ (q indexes all maturity and relative money-ness combinations $C_q = [T, k]$ for option contracts with a price record available at time t_i).

To formally establish the link between the true states and the implied states we introduce an inversion function. Denote the true state vector by $X_{t_i} = \left(y_{t_i}, v_{t_i}, \lambda_{t_i}\right)^\top$ and the implied state vector by $X_{t_i}^\theta \coloneqq \left(y_{t_i}, v_{t_i}^\theta, \lambda_{t_i}^\theta\right)^\top$, where the superscript θ emphasizes the dependence of the two implied states, $v_{t_i}^\theta$ and $\lambda_{t_i}^\theta$, on the parameter vector $\theta \in \Theta$, with Θ being a compact parameter space. Recall that $P(X_{t_i}, \theta, C)$ is the option pricing function defined in (2.17). We (implicitly) define the inverse mapping $f(p_{t_i}, \theta, C) : \mathbb{R}_+^{\overline{q}} \times \Theta \times \mathbb{R}_+^{\overline{q} \times 2} \to \mathbb{R} \times \mathbb{R}_+^2$, which recovers the latent states from option prices, through the following equation:

$$p_{t_i} = P(f(p_{t_i}, \theta, C), \theta, C). \tag{3.1}$$

Equation (3.1) defines f such that, when the latent states are recovered from a set of option prices and are then plugged back into the option pricing function P, the initial set of option

[18] We assume here that the time points are equally spaced for exposition purposes only. Unequally spaced sequences of time points can be used in the estimation routine.

prices results.

As part of the estimation procedure, the latent states are repeatedly implied, starting from an initial parameter set θ, until we arrive at a parameter set that minimizes a suitably chosen criterion function. We formalize the link between the true state vector X_{t_i} and the state vector $X_{t_i}^{\theta}$ which is obtained at the different parameter sets θ by means of the state implying function (inverse mapping) $f(\cdot)$ defined in (3.1):

$$X_{t_i}^{\theta} = f(P(X_{t_i}, \theta_0, C), \theta, C). \tag{3.2}$$

If the inversion is well-defined, which will be implied by Assumption 2 in Appendix C1, we have that $X_{t_i}^{\theta_0} = f(P(X_{t_i}, \theta_0, C), \theta_0, C) = X_{t_i}$, i.e., the true levels of the latent states can be implied when using the true parameter set θ_0.

In the large number of simulation exercises that we conducted, the levels of simulated states were always backed out with a high degree of numerical accuracy from option price panels that were generated on the basis of the respective simulated state series paths. The inversion was well-behaved at all parameter set values we experimented with, even when far from the true values of the parameters at which the state paths were simulated. Furthermore, in our empirical application, the inversion was numerically possible and well-behaved for all the sample points.

In each sample period there is a number of option price quotes available from which the latent states can be implied. Pan (2002) uses a single option contract each period to back out the level of the single latent volatility state, which suffices for her estimation. Pastorello et al. (2003) propose to introduce a pricing error with a stationary parametric distribution, which explains pricing differences in the cross-sectional dimension of the option panel and whose characteristics can be estimated together with the state dynamics. In a similar option-panel estimation context, Andersen et al. (2015a) impose weak in-fill asymptotic assumptions on the option price observation error.

Our approach is as follows. Given that the Black-Scholes implied volatility function, BSvol, is a monotonic transformation of dollar option prices, it makes no difference as to

whether the price series p_{t_i} is expressed in dollar amounts or implied volatility points. We use the latter as input in the state implying function. In our context this is further useful in reducing computational costs, as for each sample day we can build a set of implied volatilities for a standardized set of maturities and money-ness levels. So, after implying Black-Scholes volatilities from the selected sample of option prices we resort to a kernel-smoothing technique to determine a standardized set of implied volatilities. This procedure reduces the effect of microstructure noise on prices and prevents the over-weighting of in-the-money options when implying the latent state levels.[19]

We use a Gaussian kernel (see e.g., Cont and Da Fonseca (2002)) to interpolate Black-Scholes implied volatility levels[20] for option contracts with specific maturities and relative money-ness levels. At time t_i, based on the available market prices for traded contracts with characteristics C_q, the interpolated implied volatility level for a contract with new characteristics $\tilde{C} = [\tilde{T}, \tilde{k}]$ is given by:

$$\mathrm{IV}_{t_i}(\tilde{C}) = \frac{\sum_{q=1}^{\bar{q}} \mathrm{BSvol}\left(p_{t_i}(C_q)\right)\ u\left(\tilde{C} - C_q\right)}{\sum_{q=1}^{\bar{q}} u\left(\tilde{C} - C_q\right)}. \tag{3.3}$$

The smoothing function is:

$$u\left([x, y]\right) = (2\pi)^{-1} \exp\left(-x^2/2h_1\right) \exp\left(-y^2/2h_2\right), \tag{3.4}$$

with h_1 and h_2 bandwidth parameters.[21]

If option prices were observed without error, then, at each time point, two option prices would be sufficient in our setting to imply the two latent states — the stochastic volatility and the stochastic jump intensity. We impose a set of assumptions on the observation errors,

[19] The dollar level of (deep) in-the-money option prices can be several orders of magnitude larger than that of the out-of-the-money option prices, which would lead to over-weighting the former in a state implying routine.

[20] $\mathrm{BSvol}(p_{t_i}(C_q))$ denotes the function that backs out the level of volatility which, when used in the Black (1976) model for pricing options on a forward/futures contract, would equal the market price p_{t_i} for the given option contract characteristics C_q.

[21] In our empirical application, the bandwidth parameters h_1 and h_2 were chosen such that they lead to a satisfactorily smooth implied volatility surface. Adaptive bandwidth selection procedures could be considered, however the differences in the resulting standardized implied volatilities would be very small and would not lead to significant changes in the levels of the state variables backed out from these options.

detailed in Appendix C. Under these assumptions, by using in-fill asymptotic arguments, the observation errors will vanish after the kernel smoothing procedure for the implied volatility surface is applied. In practice we necessarily smooth the implied volatility surface using a finite number of option prices. Thus, we include more than two contracts in the numerical inversion procedure. In order to imply the levels of the latent state variables at each of the time points t_i, we minimize the (sum of squared) differences between the Black-Scholes implied volatilities standardized using the kernel smoothing procedure and the Black-Scholes implied volatilities obtained from the option prices generated by the model pricing approach at the assumed parameter values. Formally:

$$\begin{aligned} X_{t_i}^{\theta} &= \left(y_{t_i}, v_{t_i}^{\theta}, \lambda_{t_i}^{\theta}\right)^{\top}, \\ \text{where } \left(v_{t_i}^{\theta}, \lambda_{t_i}^{\theta}\right) &= \min_{v,\lambda} \sum_{C \in \mathcal{C}} \left(\text{IV}_{t_i}(C) - \text{BSvol}\left(P(\left[y_{t_i}, v, \lambda\right]^{\top}, \theta, C)\right)\right)^2, \end{aligned} \tag{3.5}$$

with $\mathcal{C}$ denoting a set of money-ness and maturity pairs for which interpolated implied volatilities are available at time t_i.

In our empirical application, we consider for each sample day three maturities $T \in [0.1, 0.5, 1]$ (expressed in years[22]) and a set of relative money-ness levels $k \in [0.8, 0.9, 1, 1.1, 1.2]$, such that $\mathcal{C} = [0.1, 0.5, 1] \times [0.8, 0.9, 1, 1.1, 1.2]$. After having backed out the two latent states we construct a series of observations for the full state vector which we then use for one criterion function evaluation, after which the states are backed out again at different parameter values, until the criterion function is minimized. The choice of criterion function is detailed in the next subsection.

3.2 Parameter Estimation

Exploiting both forward and option prices, the $\mathbb{P}$ and $\mathbb{Q}$ dynamics of our model will be jointly identified. We derive the conditional characteristic function under $\mathbb{P}$ in the same way as we derived the conditional characteristic function under $\mathbb{Q}$ in Proposition 1. Given the conditional characteristic function of the state vector, the conditional likelihood function of our model

[22]We follow the convention that 1 year consists of 252 trading days.

could in principle be recovered through successive applications of (numerical) multidimensional Fourier integration. However, approximating the likelihood function this way is computationally highly expensive and unsuitable for implementation in an estimation routine that requires evaluation for a large number of parameter vectors. As a more suitable alternative, we develop an approach based on GMM.

Optimal moment conditions are usually inferred from the first-order conditions of maximum likelihood, but, in our set-up, the likelihood function is not available in closed form. Singleton (2001) proposes different ways of using the conditional characteristic function for affine jump-diffusions to construct valid moment conditions which partly retain the efficiency of the optimal moment conditions stemming from the first-order conditions of maximum likelihood. We follow this route and use ϕ, the conditional characteristic function of the state vector under the physical probability measure which embeds all the ($\mathbb{P}$ and $\mathbb{Q}$) parameters of our model, to construct moment conditions.

As the state vector X_t is Markovian, we consider the following set of moment conditions based on the conditional characteristic function and its empirical counterpart:

$$h(r, s, X_{t_i}, X_{t_{i+1}}; \theta) \coloneqq m(r, X_{t_i}) \left(e^{\mathrm{i}s \cdot X_{t_{i+1}}} - \phi \left(s, X_{t_i}, \Delta; \theta \right) \right), \tag{3.6}$$

where the 3-dimensional real vector s denotes the argument at which the conditional characteristic function is evaluated and the $\mathcal{F}_{t_i}$-adapted functional $m(r, X_{t_i})$ denotes an "instrument" that may depend on a real vector r with the same dimension as the state vector. By definition of the conditional characteristic function, the set of moment conditions (3.6) constitutes a martingale difference series.

Two choices have to be made when using this set of moment conditions based on the conditional characteristic function: the choice of the arguments s and the choice of the instruments $m(r, X_{t_i})$. Regarding the former choice, Singleton (2001) shows that using a finite set of vectors $\{s_1, s_2, \dots\}$ as arguments for the conditional characteristic function such that $\{s_1, s_2, \dots\}$ have equally spaced components leads to a consistent estimator, even if such a grid of arguments contains only few distinct values for each component. For the instrument set, Sin-

gleton (2001) proposes to compute $m(r, X_{t_i})$ as an optimal instrument from the Hansen (1985) framework. In practice, although efficiency gains are obtained when the grid of arguments is refined, the inversion of the sample moment condition covariance matrix that is featured in the Hansen (1985) instrument set is numerically difficult as it often becomes near-singular in practical implementations. Carrasco et al. (2007a) investigate the efficiency of estimators based on the characteristic function by using a continuum of values for s, and show that "instrumenting" the continuum of moment conditions with a continuum of instruments of the type $m(r, X_{t_i}) = e^{\mathrm{i}r \cdot X_{t_i}}, r \in \mathbb{R}^{\dim X_{t_i}}$, leads to considerable efficiency gains in the estimation procedure.[23]

The efficient estimation procedure using a continuum of moment conditions developed in Carrasco and Florens (2000), Carrasco and Florens (2002) and Carrasco et al. (2007a) is mostly dedicated to estimation contexts in which the full state vector is observed. In our set-up, as in many other asset pricing contexts, latent state variables are present in the state vector. A potential solution in a variety of estimation problems with latent state variables, discussed in Carrasco et al. (2007a), is to replace the latent state variables by (conditionally) simulated paths in the estimation routine. An alternative solution, which we explore, is to extract information about the latent states from market prices of contingent claims linked to the conditional distribution of the state vector. We show that, under reasonable assumptions, the estimator based on a continuum of moment conditions is consistent when the latent states are backed out from option prices.

Therefore, we first imply the latent states from prices of traded derivatives and then use the implied series to construct the criterion function as if the full state vector were observable. Supposing that latent states are implied from option price data observed with a regular frequency Δ at equally spaced time points $\{t_1, t_2, \ldots, t_i, \ldots, t_n, t_{n+1}\}$ and introducing the notation $\tau \coloneqq [r, s]$, the time t_i sample moment condition in our implied state set-up is given by:

$$h_{t_i}(\tau; \theta) \coloneqq h(\tau, X^{\theta}_{t_i}, X^{\theta}_{t_{i+1}}; \theta) = e^{\mathrm{i}r \cdot X^{\theta}_{t_i}} \left(e^{\mathrm{i}s \cdot X^{\theta}_{t_{i+1}}} - \phi(s, X^{\theta}_{t_i}, \Delta; \theta) \right). \tag{3.7}$$

[23]The intuition behind this particular functional form for the instrument is that, when evaluated over a continuum of values r, it spans the partial derivative of the transition density of X_t w.r.t. the parameter set θ, which is the optimal, yet unfeasible, instrument in this set-up; see Singleton (2001) and Carrasco et al. (2007a).

We denote the sample mean of the moment condition evaluated at τ by

$$\hat{h}_n(\tau;\theta) = \frac{1}{n}\sum_{i=1}^{n} h(\tau, X^{\theta}_{t_i}, X^{\theta}_{t_{i+1}};\theta). \tag{3.8}$$

Our estimator will be based on the convergence of the moment conditions' sample mean to the unconditional expectation $\mathbb{E}^{\theta_0}\left[h_t(\tau;\theta_0)\right] = 0$ for any τ given $\theta = \theta_0$.[24]

Following Carrasco and Florens (2002), we employ a continuum of moment conditions indexed in the τ-argument by introducing a Hilbert space[25] equipped with the following inner product, for two complex-valued functions f, g, square-integrable w.r.t. a continuous density π:

$$\langle f, g\rangle \quad = \quad \int\limits_{\mathbb{R}^{\dim\tau}} f(\tau)\overline{g(\tau)}\pi(\tau)\,\mathrm{d}\tau, \tag{3.9}$$

where $\overline{g(\tau)}$ denotes the complex conjugate of $g(\tau)$ and $\pi(\tau)$ is taken to be the probability density function of the standard multivariate Gaussian density. The choice of the weighting distribution is irrelevant for the asymptotic properties of the estimator as long as it admits a continuous density function. The Gaussian density is a computationally convenient choice because of the available quadrature methods for multidimensional integration to implement the resulting estimator in practice.

Using a continuum of values for τ, we obtain a first step consistent GMM estimator by solving:[26]

$$\hat{\theta}_n^{1\text{st}} = \underset{\theta\in\Theta}{\operatorname{argmin}} \int \hat{h}_n(\tau;\theta)\overline{\hat{h}_n(\tau;\theta)}\pi(\tau)\,\mathrm{d}\tau = \underset{\theta\in\Theta}{\operatorname{argmin}} \left(\parallel \hat{h}_n(\tau;\theta) \parallel\right)^2. \tag{3.10}$$

We compute this first step estimator for our data sample by successively backing out the latent states from option prices and evaluating the integral[27] in (3.10) while varying the parameter vector θ until the integral is minimized.

Proposition 2 *Under Assumptions 1-5 (detailed in Appendix C1) the estimator* $\hat{\theta}_n^{1\text{st}}$ *converges*

[24] See Assumption 4 in Appendix C for a precise description of the operator $\mathbb{E}^{\theta_0}[\cdot]$.
[25] A formal definition of the Hilbert space is given in Assumption 3 in Appendix C.
[26] The norm $\|f\| = \langle f, f\rangle^{1/2}$.
[27] The integral can be evaluated using efficient quadrature methods.

to θ_0 in probability as $n \to \infty$.

Assumptions 1-5 ensure consistency in our (generic) estimation set-up in which two of the latent states are backed out from option prices. In Appendix C we discuss the validity of these assumptions for the specific model with self-exciting jumps which we consider in our application. We note, however, that the proposition (and, similarly, Proposition 3 below) pertains more broadly to estimating other models (e.g., other affine jump-diffusions) that meet the imposed assumptions.

In the spirit of Carrasco et al. (2007a), we also compute a second step GMM estimator which uses the first step estimator as an input to calculate the covariance operator, K, applied to our moment conditions and given by:

$$Kf(\tau_1) = \int k(\tau_1, \tau_2) f(\tau_2) \pi(\tau_2) \, \mathrm{d}\tau_2, \tag{3.11}$$

$$\text{with} \quad k(\tau_1, \tau_2) = \mathbb{E}^{\theta_0} \left[h_t(\tau_1; \theta_0) \overline{h_t(\tau_2; \theta_0)} \right]. \tag{3.12}$$

This is akin to finding an estimator for the optimal weighting matrix in the second step of GMM estimation with a finite number of moment conditions, and now adapted to the context in which a continuum of moment conditions is employed. When the full Markovian state vector is observed, the inverse of the covariance operator is known to yield an estimator with minimal variance, which, subject to further conditions, attains the Cramer Rao efficiency bound (Carrasco et al., 2007a). In the current set-up, the estimator for the covariance operator relies instead on the latent states implied from option prices. This introduces a non-explicit dependence of the covariance operator on the parameter vector which makes analytic derivations of the efficiency gains resulting from a second step estimator infeasible. Nevertheless, the second step estimator is expected to lead to an efficiency improvement. This second step estimator is defined as:

$$\hat{\theta}_n^{2\text{nd}} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left(\| \, (K_n^{\alpha_n})^{-1/2} \hat{h}_n(\tau; \theta) \, \| \right)^2. \tag{3.13}$$

Here, the superscript α_n denotes the regularization of the inverse of the covariance operator K and of its empirical counterpart K_n, which is needed to ensure a stable inverse; see Appendix

C for further details.[28] The empirical counterpart of K given by K_n is obtained by estimating the kernel k in (3.12) as a sample average evaluated using the first step estimator.

Proposition 3 *Under Assumptions 1-7 (detailed in Appendix C1) the estimator* $\hat{\theta}_n^{2\text{nd}}$ *converges to* θ_0 *in probability as* $n \to \infty$.

3.3 Monte Carlo Evidence and Practical Considerations

To assess the finite sample performance of the proposed estimation approach we conduct an extensive Monte Carlo exercise. Below we summarize the main results and we provide a separate online appendix with additional results on identification of the latent states and the model parameters and with computational details related to our estimation routine.[29]

We simulate state vector series from the stochastic volatility with Hawkes self-exciting jumps model (labeled "SVHJ") and then price options using the conditional characteristic function based approach presented in Section 2 to create panels of option prices which serve as data input for estimation. Specifically, using realistic parameter values that mimic those obtained in our empirical analysis presented later, we simulate state vector levels by discretizing[30] their continuous-time dynamics. Each simulated sample contains 9 option prices (maturities 0.1, 0.5 and 1 year and money-ness levels of 95%, 100%, and 105%) for each of 500 weekly observations. Therefore, in the T dimension, the simulated option panel contains 500 observations, while the cross-sectional dimension is constant for each t and set to 9 option prices. We run the estimation for 100 simulated option panels, thus obtaining 100 sets of parameter estimates. This number of samples is perhaps small compared to the typical number of replications used in Monte Carlo exercises, but, given the high computational costs, it provides a feasible implementation which gives a reasonable indication about the small sample behavior of our estimator. Table 1 presents the parameter values used in the simulation, the corresponding Monte Carlo means, sample standard deviations, and selected quantiles of the estimator's

[28] An overview of ill-posed inversion problems and regularization schemes is given in Carrasco et al. (2007b). A simplified computation procedure for the second step estimator involving the inverse of the covariance operator using matrix expressions is detailed in Proposition 3.4 of Carrasco et al. (2007a) and can be adapted to our setting.

[29] Available from the authors' webpages; see `http://www.rogerlaeven.com`.

[30] We adopt Euler discretization at a very fine time grid; see e.g., Lord et al. (2010) for details.

empirical distribution.

The Monte Carlo results are indicative of a good finite sample performance. The parameters characterizing the stochastic jump intensity ($\kappa_\lambda, \overline{\lambda}$ and δ), which are of particular interest, are pinned down with good precision. Furthermore, the parameters related to the stochastic volatility process ($\kappa_v, \overline{v}, \sigma_v$ and ρ) are identified with high precision. Estimates for the log-normal mean jump size under the physical probability measure, $\mu_j^{\mathbb{P}}$, are less precise, a likely consequence of the fact that the identification of $\mu_j^{\mathbb{P}}$ and η comes from the time series of log-forward returns. Therefore, this warrants the caveat that while risk premium inference based on finite sample estimates is informative, it can lack precision when using option panels with a relatively small time-series dimension.

[Table 1]

As a matter of fact, besides the appealing theoretical characteristics of the estimator based on a continuum of moments, its development in the current set-up was also motivated by identification concerns and practical considerations we faced when using alternative estimation procedures. We have implemented several other "GMM-type" estimators based on the conditional characteristic function, such as those suggested by Singleton (2001) and Pan (2002), which can in principle be extended to apply to the current set-up. However, while computing the instrument sets needed for these alternative estimation procedures, infeasible computational costs arise, because in our set-up we necessarily rely on numerical option pricing for every criterion function evaluation. With such alternative estimators, we also encountered some finite sample identification problems particular to our model. They are mainly due to the different scale which the stochastic jump intensity can have compared to the other state variables.[31] These issues are remedied when using the estimation procedure proposed in this paper.

Because the latent state implying step of the estimation procedure relies on a numerical option pricing routine and the estimator of the covariance operator relies on the latent states implied from option prices, which implicitly depend on the first step estimator $\hat{\theta}_n^{1\text{st}}$, analytic

[31] Re-scaling the state variables improves finite sample performance. However, in studies based on simulated data samples we found that, not knowing the sample maximum of the implied stochastic jump intensity levels ex ante, made the re-scaling procedure an ad hoc solution for achieving satisfactory finite sample performance.

results about the asymptotic variance and efficiency of the resulting second step estimator in the context of the SVHJ model cannot be derived.[32] Therefore, we need to find an approximation device to obtain (asymptotic) standard errors of parameter estimates for our empirical analysis. To this avail, we resort to computing a numerical gradient-based approximation of GMM-type standard errors using the second step criterion function.

4 Empirical Application

4.1 Data Description

We apply our estimation procedure to options data on the S&P 500 index from the Option Metrics database. The data-set we use contains closing bid and ask prices for all Chicago Board Options Exchange traded European style option contracts (both put and call contracts). The data-set spans the period from the first calendar week of January 1996 to the last calendar week of August 2013. We split the data-set into a first in-sample part, covering the period from January 1996 to December 2009, and a second out-of-sample part, covering the period from January 2010 to August 2013; see Appendix D. We retain an out-of-sample period to assess whether model parameters estimated using data from the in-sample period would give satisfactory results when used for state backing out and option pricing during the out-of-sample period. The long time span of our in-sample data-set covers some of the well-known turbulent periods around the dot-com bubble burst in 2000 as well as the bankruptcy of Lehman Brothers, which occurred in September 2008, and its aftermath. The in-sample period includes many significant drops in the S&P 500 index level which happened in rapid succession around those time periods and could easily be reconciled with the self-exciting jump framework we use. It is worth to mention that our model is designed to capture both jump clusters and regular day-to-day variations, so we expect that the model achieves a good fit during non-turbulent periods as well as turbulent ones.

[32] A formal justification for the second step estimator by expressing the asymptotic efficiency gain for the estimator can therefore not be made. We nevertheless argue that it is reasonable to expect that a weighting of the moment conditions based on the inverse covariance operator evaluated at a consistent first step estimate leads to an efficiency gain, even despite the fact that the latent states used in the computation of the covariance operator are implied from option price observations at the first step estimator $\hat{\theta}_n^{1_{\text{st}}}$, thus introducing an additional implicit dependence on the first step estimator.

Because the data-set contains daily closing prices for a large number of option contracts, it would be computationally over-demanding to back out the implied states, compute model option price counterparts, and conduct inference using all the option prices recorded in the sample. Instead we follow inter alia Pan (2002) and Johannes et al. (2009) and use weekly options data. The day of the week selected to include in our sample is Wednesday; this is a typical choice in the literature. Using panels of options data to estimate parameters of continuous-time stochastic processes poses some specific challenges which we describe henceforth.

Given that the data-set covers a long time period, which likely spans more than a single business cycle, one must find a way to account for time variation in interest rates and time variation in dividend yields. This constitutes an important modeling step because the interest rate and dividend yield levels are parameters which appear in the drift specification of the log-index return process together with the time-varying jump and volatility risk premiums, and hence influence the reliability of the inference conducted on the latter. Modeling the log-forward returns instead of log-index returns alleviates this issue. One could think of using daily closing prices for traded futures contracts to obtain an approximation for the log-forward return, but, due to the different venues in which futures contracts and options trade, one cannot fully rely on the closing bid and ask prices for futures contracts. We take the route of Aït-Sahalia and Lo (1998) and back out the forward rate from the put-call parity for European options, thereby circumventing this problem.[33]

Our estimation procedure also makes use of a second data-set which provides interest rate levels for maturities traded on the respective sample days. To obtain other points on the yield curve, based on the available yields, we apply an interpolation procedure for each sample day.

4.2 Data Selection Procedure

We prepare our data-set for estimation by selecting for each week in the sample the Wednesday closing bid and ask prices for call and put options and by using the midpoint between the closing

[33] Other approaches to account for these options panel specific issues exist in the literature. Pan (2002) models the dividend yield and interest rate levels separately, each as an affine diffusion and then uses them together with the log-index returns in the option pricing routine. As we did not have access to synchronized recordings of index and option price data we did not follow this route.

bid and ask prices to serve as an approximation for each option's trading price. We discard some of the observations by retaining in the sample only the option contracts which meet all of the following criteria: maturity of the contract in trading days is $\in$ [20 days, 300 days], the mid-price > 0.1 (10 USD cents[34]), and money-ness $\in [50\%, 150\%]$. We apply the same criteria for put and call options.

From the selected options we build a standardized data-set. To do so, we imply, for each sample day, t, and for each maturity, T, that is available in that sample day, the forward price $F_{t,T}$ for the forward contract with maturity T using the (closest to) at-the-money pair of call and put prices in the put-call parity for options written on a forward contract:

$$\text{Call}_t + Ke^{-r_{t,T}(T-t)} = \text{Put}_t + F_{t,T}e^{-r_{t,T}(T-t)}, \tag{4.1}$$

where $r_{t,T}$ is the continuously compounded risk-free interest rate at time t with maturity T.

To complete the standardized data-set, we again use the put-call parity (4.1) to determine in-the-money option prices using their out-of-the-money counterparts and the previously determined forward price, as in Aït-Sahalia and Lo (1998). Thus, we re-calculate the in-the-money call prices using the out-of-the-money put prices together with the corresponding maturity forward price level determined from at-the-money options. By doing so, we replace the original in-the-money option prices, which are likely imprecise due to the typical illiquidity of these contracts, with the recalculated ones which should be a more precise approximation of the "true" price of (deep) in-the-money options. This builds a range of option prices for both in-the-money and out-of-the-money contracts for each sample day and each maturity using the majority of available put and call contracts in the original data-set.

Finally, based on this adjusted panel of option prices, we use the kernel smoothing procedure discussed in Section 3.1 to obtain the standardized sets of Black-Scholes implied volatilities, which constitute the data input for the latent state implying step in the estimation procedure.

[34]This is the minimum tick on the CBOE during our sample period.

4.3 Parameter Estimates and Implied States

Table 2 presents our GMM estimates for the model parameters and standard errors (in parentheses) using the in-sample weekly (Wednesday's) options data on the S&P 500 index spanning January 1996 to December 2009.

[Table 2]

All parameter estimates are statistically significant at the conventional 5% significance level, with the exception of η, the parameter governing the diffusive risk premium, which is significant only at the 10% level. Note further that $\overline{v}$ is the long run average "variance" and that the long run average volatility as deduced from Table 2 is $\sqrt{\overline{v}} \approx 10.5\%$.

Our model features two sources that generate "variation" in returns: one stemming from the stochastic volatility process and the other stemming from the self-exciting jump process. The time series of the two state variables can be backed out from option prices using the estimates in the table above. Figure 1 plots the time series of the two implied states, the stochastic volatility of the asset $\sqrt{v_t}$ (with the weekly *returns* for the S&P 500 index in the background) and the stochastic jump intensity λ_t (with the *level* of the S&P 500 index in the background). Our estimation procedure does not confine the implied jump intensity process to behave like the intensity of a self-exciting process. In particular, the implied state in week t_i is unrelated to the implied state in week t_{i+1}.[35] It is remarkable that the time series of implied jump intensities resembles a typical simulated sample path from a self-exciting process.

The implied jump intensity quickly increases around the past dates of turbulent events such as the 1998 Russian Ruble crisis, the dot-com bubble in the 2000's, and the global financial crisis which began in late 2008, suggesting that these events are naturally accommodated in the self-exciting jump framework. The estimates $\hat{\kappa}_\lambda = 18.16$ and $\hat{\delta} = 16.62$ are indicative of an intensity process with relatively persistent jumps. Given these values for $\hat{\kappa}_\lambda$ and $\hat{\delta}$, the half-life of a shock to the intensity generated by an asset price jump is approximately two weeks.[36] The unconditional expected jump intensity, $\mathbb{E}[\lambda] = \kappa_\lambda \overline{\lambda}/(\kappa_\lambda - \delta)$, is estimated at 3.8, meaning that unconditionally one would expect a jump in the S&P 500 index to occur about every quarter.

[35] The only explicit link between consecutively implied latent states is the feature that the states are implied using the same option pricing model with the estimated model parameter values.

[36] Assuming no further jumps occur in the meanwhile.

4.4 Model-Implied Volatility Surfaces

Our model is quite flexible in accommodating the various profiles that the implied volatility surface can have over time.[37] This flexibility stems from the two stochastic state variables that vary continuously and jointly determine changes in the shape of the implied volatility surface. The stochastic jump intensity and the stochastic volatility each play their own role such that, as we will see, a good fit is achieved without having to periodically re-estimate model parameters.[38]

Option pricing frameworks featuring a jump component in the return process that follows the prototypical Poisson process or a doubly stochastic Poisson process generate a steep skew in the model-implied volatility profile for short maturity options. Thus, the jump component aides these models in fitting the short maturity skew of the implied volatility surface that is empirically observed. However, in these frameworks, the dynamics of the implied volatility surface are not affected by jump events themselves, i.e., the model-implied volatility surface skew is invariant to jump occurrences. By contrast, the jumps in our model have an impact on the temporal dynamics of the implied volatility skew. This is due to the dependence of the option price upon the level of the jump intensity and given that, in the self-exciting model, jumps feedback into the intensity process, jumps consequently feedback into option prices and model-implied volatility surfaces.

Figure 2 plots the implied volatility surface dynamics over time after a jump occurs. More specifically, it plots the volatility surface sequentially over time, starting from an initial time point t for which we impose $\mathrm{d}N_t = 1$ and assuming that no further jumps occur after the initial one at time t. The plot isolates the impact of jumps on option prices, as the volatility state level, which also plays a role in determining the shape and dynamics of the model-implied volatility surface, is kept constant at all time points, being set equal to its long run mean $\overline{v}$.

[Figure 2]

Upon the occurrence of a jump, according to our parameter estimates, the instantaneous

[37] Studying the dynamics of the instantaneous profile of the implied volatility surface is a way to analyze the evolution of option prices over time.

[38] In option pricing models that rely on only a single stochastic state variable or do not feature feedback from shocks to state variables, the parameter sets usually have to be re-estimated to cope with inherent changes in market regimes and accommodate the associated dynamics of the implied volatility surface.

level of the jump intensity becomes $\lambda_{t_{\text{jump}}} \coloneqq \overline{\lambda} + \delta$, which is estimated at 16.95. The sudden increase in the level of the jump intensity due to an asset price jump generates an upward level shift of the stochastic volatility surface along with a pronounced change in skew for shorter maturities, the latter being the region of the implied volatility surface that is most sensitive to the level of the jump intensity state. In the absence of subsequent jumps, the surface shifts downwards due to the exponential decay of the jump impact on the intensity level, while the skew corresponding to short maturity options also adjusts.

4.5 Model Fit and Option Pricing Errors

Compared to alternative models with standard stochastic volatility and jump specifications, the self-excitation feature of our jump process is expected to induce a decrease in option pricing errors, especially when these are tracked over longer periods of time spanning multiple episodes of asset price crashes. We investigate model fit by comparing the option pricing performances of several stochastic volatility models with jumps. The models we consider are: stochastic volatility with prototypical Poisson jumps (labeled "SVJ"), stochastic volatility with volatility driven jump intensity (labeled "SVVJ"), and stochastic volatility with Hawkes self-exciting jumps (our SVHJ model). The exact specifications of the SVJ and SVVJ models under the physical measure $\mathbb{P}$ and the risk neutral measure $\mathbb{Q}$ are presented in Appendix E. The SVVJ model was adopted by Pan (2002). All three models share the same stochastic volatility process (but with potentially different parameter values), the same assumptions for constructing the candidate pricing kernel, and are estimated using the same procedure provided by our implied-state GMM with a continuum of moments. Thus, the focus of the comparison is on the impact of the jump component specification on overall option pricing performance.

The SVJ and SVVJ models feature only one latent state, the stochastic volatility. By backing out the stochastic volatility levels from a panel of actual derivatives prices, we construct a time series of (implied) state vectors, which we further use to compute a continuum of moment conditions derived from the conditional characteristic functions of these benchmark models. The data sample we employ for the estimation of the benchmark models is exactly the same as the one we use for the estimation of the SVHJ model, i.e., S&P 500 option price data covering

the time span from January 1996 to December 2009.[39]

To evaluate option pricing performance we compute typical option pricing error metrics:[40] the average and root mean squared differences between market- (observed) implied volatilities and model- (generated) implied volatilities during the full sample period (January 1996 to August 2013). At every sample time point, the latent state(s) for each model are backed out[41] from the sub-sample of derivatives contracts used for estimation in the case of the in-sample period and, similarly, from the sub-sample of derivatives contracts used for assessment for the out-of-sample period. With the latent state(s) backed out, the implied volatility surface at the remaining grid points for which, at that point in time, market quotes are available, is computed using model-implied prices.[42] The pricing differences (market vs. model) for each of the three models are then compared to reveal which model is better at fitting the implied volatility surface. Each model's pricing performance can be evaluated by tracking these pricing differences for different contract maturities and money-ness levels over time.

[Figure 3]

[Figure 4]

Figures 3 and 4 plot the mean and root mean squared pricing errors, respectively, expressed in implied volatility percentage points. The SVHJ model with self-exciting jump intensity clearly leads to the smallest pricing errors, albeit that short maturity deep out-of-the-money options are still on average slightly underpriced, even by the SVHJ model. The SVVJ model ranks second in terms of pricing performance, also underpricing short maturity, deep out-of-the-money options, whilst overpricing the remaining maturity and money-ness combinations. The poor performance of the SVJ model can be indicative of model misspecification, given that the model parameters are estimated from a long time-series sample spanning different market regimes. The need for an additional stochastic factor besides the stochastic volatility is evident in this set-up and among the two stochastic intensity models, the SVHJ results in

[39] The parameter estimates for the SVJ and SVVJ models are presented in Appendix E.

[40] A typical option pricing benchmarking exercise can be found in Bakshi et al. (1997).

[41] Using the parameter estimates previously obtained for each of the models via the implied-state GMM procedure with a continuum of moments.

[42] In other words, each of the three models is used as an interpolation tool for the implied volatility surface: assuming that at some point in time, a few grid points of the implied volatility surface are observed, the latent state(s) is (are) backed out, and employed to price "new" derivative contracts, corresponding to additional grid points on the surface. These new prices are compared with the actual market prices recorded.

smaller option pricing errors achieving a better fit for market-implied volatility surfaces.

5 Risk Premiums and Investor Fear Gauges

5.1 Equity Risk Premium

The candidate pricing kernel (2.7)–(2.9) induces two time-varying risk premiums: a diffusive risk premium and a jump risk premium. The sum of the two premiums represents the total compensation required by investors due to the uncertainty about the future values of the forward price and is typically referred to as the equity risk premium. Specifically, following Bollerslev and Todorov (2011), we define the equity risk premium, ERP, as:

$$\mathrm{ERP}_t = \frac{1}{T-t}\left(\mathbb{E}_t\left[\frac{F_T - F_t}{F_t}\right] - \mathbb{E}_t^{\mathbb{Q}}\left[\frac{F_T - F_t}{F_t}\right]\right), \tag{5.1}$$

where $\mathbb{E}_t$ and $\mathbb{E}_t^{\mathbb{Q}}$ are shorthand notations for the $\mathcal{F}_t$-conditional expectations. A natural decomposition of the above (generic) expression for the equity risk premium occurs by separating the contribution of the two sources of risk specified in our SVHJ model, that is:

$$\mathrm{ERP}_t = \mathrm{ERP}_t^{\mathrm{dif}} + \mathrm{ERP}_t^{\mathrm{jump}}, \tag{5.2}$$

$$\text{where} \quad \mathrm{ERP}_t^{\mathrm{dif}} = \frac{1}{T-t}\int_t^T \eta \mathbb{E}_t\left[v_s\right] \mathrm{d}s, \tag{5.3}$$

$$\text{and} \quad \mathrm{ERP}_t^{\mathrm{jump}} = \frac{1}{T-t}\int_t^T (\mu - \mu^*)\mathbb{E}_t\left[\lambda_s\right] \mathrm{d}s. \tag{5.4}$$

Fully explicit expressions for $\mathrm{ERP}_t, \mathrm{ERP}_t^{\mathrm{dif}}$ and $\mathrm{ERP}_t^{\mathrm{jump}}$ are provided in Appendix F.

Evaluated at the long run mean volatility level for the S&P 500 index, by using the estimated value of $\overline{v}$ equal to 0.011, the average long run diffusive risk premium would be approximately 2.61% per year. However, we note that the actual $\mathrm{ERP}_t^{\mathrm{dif}}$ is time-varying and determined by the fluctuations in the stochastic volatility state v_t.

The jump risk premium is dictated by the difference between the relative jump sizes of

index price jumps under the two equivalent measures, and is given by $(\mu - \mu^*)\lambda_t$ per unit of time. The case for a richer specification of the jump risk premium can be made, but would lead to challenging identification issues. Despite the standard affine specification of the jump risk premium in our model, the patterns of the jump risk premium over time are argued to be realistic thanks to the self-exciting jump framework used. The jump risk premium increases abruptly in times of turmoil, an empirical finding which is consistent with the intuition that investors exhibit increased fear of future jumps after jumps occur and hence they will only be willing to bear this risk if compensated. Furthermore, in times of crisis, the jump risk premium swamps the diffusive risk premium.

This pattern can be observed in Figure 5 which plots the diffusive risk premium, the jump risk premium, and the total equity risk premium over time, where the in- and out-of-sample distinctions follow the same conventions as in the previous section.

[Figure 5]

5.2 Risk Premiums and Option Prices

The risk premiums affect the joint fit of forward returns and option prices. While the diffusive risk premium only affects the drift of the forward return process, the jump risk premium (which is jump-size related and not jump-timing related) impacts both the drift of the forward return process and the jump size distribution, as the mean jump size under the physical measure is different from that under the risk neutral measure.

To analyze the impact of the jump risk premium on option prices we compute "hypothetical option prices" obtained by replacing $\mu_j^{\mathbb{Q}}$ by $\mu_j^{\mathbb{P}}$ throughout (*ceteris paribus*) and using our parameter estimates from Section 4. This "turns off" the impact of the jump risk premium on the jump size distribution under $\mathbb{Q}$ while maintaining martingale dynamics,[43] thus allowing to depict its impact on option prices; cf. Pan (2002). We compute a series of European call option prices with the state variables set equal to their estimated long run means, i.e., $v_t \equiv \overline{v}$ and $\lambda_t \equiv \overline{\lambda}$.

[43] In other words, the hypothetical log-forward returns y_t retain martingale dynamics under $\mathbb{Q}$, but with the jump component $\mathrm{d}J^{\mathbb{Q}}$ and its compensator set equal to their counterparts under $\mathbb{P}$.

[Figure 6]

Figure 6 shows the intricate significant impact of the jump risk premium on the implied volatility surface.[44] In-the-money short time-to-maturity call options, the put-call parity counterparts of which are out-of-the-money short time-to-maturity put options, are more valuable when jump size risk is compensated.[45] This suggests that the jump risk premium in the model plays an important role in explaining the empirical patterns of out-of-the-money put options. Our empirical evidence of a significant jump risk premium is broadly in line with the findings in Broadie et al. (2009), who argue that a significant jump risk premium in stochastic volatility models with jumps likely serves as an explanation for what otherwise seem to be abnormal returns for put option based trading strategies such as put spreads, at-the-money straddles or delta-hedged puts.

5.3 Variance Risk Premium as an Investor Fear Gauge

The CBOE volatility index (VIX, CBOE (2009)) is often interpreted as a measure of investors' fear of equity market crashes. Some studies (e.g., Whaley (2009), Bollerslev and Todorov (2011), Schneider and Trojani (2015)) have emphasized that although the VIX certainly contains information about market sentiment it is nevertheless a "noisy" measure of investor fear.

In the context of non-parametric approaches, Bollerslev and Todorov (2011) and Bollerslev et al. (2015) have proposed alternative measures to gauge investor fear of (negative) downward movements. We develop a similar fear gauge in the parametric context of the SVHJ model by defining it to be the time-varying difference between the expected future (integrated) quadratic variations under the equivalent probability measures $\mathbb{P}$ and $\mathbb{Q}$. Specifically, we define

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

[44] Relative price contributions are tabulated in Appendix G.

[45] The log-return process is a martingale, so changing the jump size mean does not impact option prices through first-order moments of the process. However, a change in the jump size mean does impact the volatility and higher-order moments of the return process, and option prices depend crucially on the volatility and higher-order moments of returns. In our case, the jump risk premium translates mainly into higher implied volatilities, hence explaining the effect seen on the right-hand side of the surface. At the same time, the jump risk premium also means more negative skewness. Skewness in the return distribution affects the shape of the smirk by "steepening" it and the increased skewness is to some degree reflected on the left-hand side of the surface.

where, under the SVHJ model,

$$\mathrm{QV}_{[t,T]} = \int_t^T v_s \, \mathrm{d}s + \int_t^T Z_s^2 \, \mathrm{d}N_s.$$

This difference is often referred to in the literature as the variance risk premium, hence VRP, being analogous to the difference between the strike level of a variance swap and the expected realized variance until the swap's maturity.

In the SVHJ model this difference stems entirely from the jump risk premium because

$$\mathrm{VRP}_t = \frac{1}{T-t} \left(\mathbb{E}\left[Z_t^2 \right] - \mathbb{E}^{\mathbb{Q}}\left[Z_t^2 \right] \right) \int_t^T \mathbb{E}_t \left[\lambda_s \right] \mathrm{d}s. \tag{5.5}$$

The jump risk premium captures the jump related compensation investors would demand for holding the market index. Consequently, the variance risk premium can naturally be interpreted as a time-varying gauge of investor crash fears.

While other fear measures, e.g., Bollerslev and Todorov (2011), can perhaps claim more robustness by relying on non-parametric procedures, the computation of the fear gauge we propose in the context of the SVHJ model is in principle straightforward given our estimation procedure.[46] It provides a parsimonious parametric approach with a clear interpretation of the parameters involved. Furthermore, we find that even with the simple affine specification of the jump risk premium we adopt, the self-exciting jump model captures most of the dynamics of the variance risk premium found in non-parametric studies. This can be observed from the variance risk premium plot in Figure 7, in which the SVHJ parametric fear gauge largely overlays the non-parametric variance risk premium from Bollerslev and Todorov (2011). Both measures are expectations computed over 1-month ahead intervals, so as to be comparable. We emphasize, however, that the two measures are computed using different estimation approaches and different data selection procedures, and therefore no exact interpretation for their differences can be given.

[Figure 7]

[46] A fully analytic expression for VRP_t is provided in Appendix F.

As depicted in Figure 7, past major market disruptions are clearly picked up by the fear gauge. The October 1997 "mini-crash", the LTCM debacle of September 1998 (triggered by the Russian sovereign default), the event of September 11, 2001, and the "dot-com" bubble of 2002 all show up as spikes in the index. The investor fear gauge ramps up the most during the global financial turmoil period of 2008-2009. Subsequent smaller increases correspond to the times during which the 2010 Euro sovereign debt crisis and the 2011 United States debt ceiling crisis unfolded.

6 Conclusions

We propose a parametric model for asset returns and option pricing based on self-exciting jump processes to accommodate clustered price jumps. We develop a consistent estimation methodology for our model based on a continuum of moment conditions which are evaluated by implying latent stochastic states — stochastic volatility and stochastic jump intensity — from a panel of option prices. Our resulting estimator has good finite sample properties. Model estimates for the S&P 500 index options show evidence in favor of self-excitation in jumps. The model provides an improved fit for option prices and further allows for an analysis of the risk premiums embedded in option prices and the development of a measure of investor crash fears.

Acknowledgements. We are very grateful to Yacine Aït-Sahalia, Federico Bandi, Christian Brownlees, Patrick Gagliardini, Frank Kleibergen, Eric Renault, Enrique Sentana, Peter Spreij, Fabio Trojani, Grigory Vilkov and to conference and seminar participants at the 8^{th} Annual SoFiE Conference, 2015 European Meeting of Statisticians, 2015 SoFiE Financial Econometrics Spring School, Dynstoch 2015, 14^{th} Winter School on Mathematical Finance, 2015 Netherlands Econometric Study Group meeting, the University of Amsterdam and the Tinbergen Institute for helpful discussions and comments. This research was funded in part by the Netherlands Organization for Scientific Research (NWO) under grant NWO Vidi 2009 (Laeven).

Appendix A

Simulated Paths from a Hawkes Jump Process

[Figure 8]

Appendix B

Candidate Pricing Kernel

Let $\tilde{S}_t = \xi_t S_t \exp\left(\int_0^t \zeta_s \, \mathrm{d}s\right)$ denote the gain earned by holding one unit of stock (index) and continuously reinvesting all dividend gains ζ_t in stock (index) units deflated by the state price density ξ_t defined in (2.7)–(2.9). Let $\tilde{B}_t = \xi_t \exp\left(\int_0^t r_s \, \mathrm{d}s\right)$ denote the gain generated by holding a dollar in the bank account deflated by the same state price density ξ_t.

The state price density process ξ_t in (2.7)–(2.9) ensures (with the relevant parameter restrictions imposed) that the deflated stock price process $\mathrm{d}\tilde{S}$ and the deflated money market account process $\mathrm{d}\tilde{B}$ are local martingales. This can be shown by applying Itô's formula for these two processes, as follows:[47]

$$\mathrm{d}\tilde{S}_t = \mathrm{d}\left(\xi_t S_t \exp\left(\int_0^t \zeta_s \, \mathrm{d}s\right)\right) = \left(\sqrt{v_t} - \Gamma_{1,t}\right) \tilde{S}_t \, \mathrm{d}W_t^{(1),\mathbb{P}} - \Gamma_{2,t} \tilde{S}_t \, \mathrm{d}W_t^{(2),\mathbb{P}} + \left(\exp\left(Z_t^* + Z_t\right) - 1\right) \tilde{S}_{t_-} \, \mathrm{d}N_t - \lambda_t \mu^* \tilde{S}_t \, \mathrm{d}t, \quad \text{(B.1)}$$

$$\mathrm{d}\tilde{B}_t = \mathrm{d}\left(\xi_t \exp\left(\int_0^t r_s \, \mathrm{d}s\right)\right) = -\Gamma_{1,t} \tilde{B}_t \, \mathrm{d}W_t^{(1),\mathbb{P}} - \Gamma_{t,2} \tilde{B}_t \, \mathrm{d}W_t^{(2),\mathbb{P}} + \left(\exp(Z_t^*) - 1\right) \tilde{B}_{t_-} \, \mathrm{d}N_t. \quad \text{(B.2)}$$

Here the marginal distribution of Z_t^*, the time t random jump size of log-kernel jumps, is normal with mean μ_j^* and standard deviation σ_j^*, for which the restriction $\mu_j^* + \frac{1}{2}\left(\sigma_j^*\right)^2 = 0$ is imposed. This restriction entails that the jump related risk premium is due to jump size risk only, i.e., the candidate pricing kernel attaches no premium to jump timing. The time t random jump size of log-kernel jumps is allowed to be correlated with the time t random

[47] We assume that the dividend process $\zeta_t, t \geq 0$, is a non-negative deterministic process.

jump size of log-index returns, Z_t. We denote the corresponding correlation coefficient by the parameter ρ^*. Then, it follows that:

$$\begin{aligned}\mathbb{E}\left[\exp\left(Z_t^* + Z_t\right) - 1\right] &= \exp\left(\mu_j^* + \frac{1}{2}(\sigma_j^*)^2 + \mu_j^{\mathbb{P}} + \sigma_j^*\sigma_j\rho + \frac{1}{2}\sigma_j^2\right) - 1 \\ &= \exp\left(\mu_j^{\mathbb{P}} + \sigma_j^*\sigma_j\rho^* + \frac{1}{2}\sigma_j^2\right) - 1 \\ &= \exp\left(\mu_j^{\mathbb{Q}} + \frac{1}{2}\sigma_j^2\right) - 1 \\ &= \mu^*,\end{aligned}$$

with $\mu_j^{\mathbb{Q}} = \mu_j^{\mathbb{P}} + \sigma_j\sigma_j^*\rho^*$, and

$$\mathbb{E}\left[\exp\left(Z_t^*\right) - 1\right] = \exp\left(\mu_j^* + \frac{1}{2}(\sigma_j^*)^2\right) - 1 = 0.$$

These results imply that the dynamics in (B.1) and (B.2) indeed describe two local martingales, as the terms involving jumps are compensated.

Finally, note that if we define the density process $\xi_t \exp\left(\int_0^t r_s \, \mathrm{d}s\right)$ we obtain a local martingale process which can be shown to be a martingale almost surely. The density process can then be used to generate an equivalent martingale measure $\mathbb{Q}$, under which

$$W_t^{(1),\mathbb{Q}} = W_t^{(1),\mathbb{P}} + \int_0^t \Gamma_{1,s} \, \mathrm{d}s \qquad \text{and} \qquad W_t^{(2),\mathbb{Q}} = W_t^{(2),\mathbb{P}} + \int_0^t \Gamma_{2,s} \, \mathrm{d}s. \tag{B.3}$$

Using (B.3) and taking into account the change in the mean expected jump size, the dynamics of the state variables under the risk neutral (pricing) measure $\mathbb{Q}$ are those described in (2.11)–(2.13).

Appendix C

This appendix first discusses technical assumptions which we impose on the data generating process and the associated observation errors and on the estimation procedure, and next proves that under these assumptions the estimator we develop is consistent.

More specifically, in Appendix C0, we impose a prerequisite set of assumptions on the observation errors resulting from the fact that we can only observe "noisy" option price panels for which we use the mid-price as a proxy for the "true" (efficient market) option price. Appendix C1 describes the assumptions we impose to establish consistency of our estimator, given the structure of the observations. These assumptions in principle resemble the typical assumptions imposed for consistency of GMM estimators in set-ups in which a continuum of moments is employed, e.g., Carrasco and Florens (2000), Carrasco and Florens (2002), and particularly Carrasco et al. (2007a). The crucial difference between our approach and these earlier works, however, is that we use option-implied states to fill in for the otherwise unobservable elements contained in the Markovian state vector. Thus, the moment conditions in our set-up are intricate functions of the parameter set, which makes our set-up different from, and in this respect more involved than, the original aforementioned frameworks. Under the set of assumptions that we formulate in our implied-state set-up, consistency of our estimator can be shown. We provide the proof of this result in Appendices C2 and C3.

For the particular SVHJ option pricing model that we propose, the latent states cannot be expressed as fully analytic functions of the parameter vector. Therefore, we discuss in Appendix C4 the implications this has on the validity of the assumptions imposed in Appendix C1. The discussion is provided in the context of the SVHJ model. However, it applies more generally to option pricing models with latent stochastic states for which a fully analytic expression of the characteristic function is not available.

C0. Assumptions for the In-Fill Asymptotic Behavior of Observation Errors

We consider in-fill asymptotics to deal with observation errors: we assume that, as the cross-sectional dimension of the panel of option quotes increases, the differences between the Black-Scholes implied volatilities corresponding to the true (efficient market) option prices and the Black-Scholes volatilities implied from the mid-price quotes we have available vanish asymptotically after the kernel smoothing procedure is applied.

Formally, as in Andersen et al. (2015a), we assume that the implied volatility levels are observed with additive noise. We further assume that the implied volatility observation errors

have zero mean:[48,49]

Assumption 0:

(i) $\text{BSVol}\left(p_t(C)\right) = \text{BSVol}^0\left(p_t(C)\right) + \varepsilon_t(C)$, i.e., the observation error, ε_t, is additive.

(ii) $\mathbb{E}\left[\varepsilon_t(C)\right] = 0, \mathbb{E}\left[\varepsilon_t(C)\right]^2 < \infty, \ \forall t$ and $\forall C$, i.e., observation errors do not translate into a systematic bias for the implied volatility levels and have finite variance.

The two conditions stated in Assumption 0 imply that a Law of Large Numbers holds (see e.g., Taylor (1978)) such that:

$$\frac{\sum_{q=1}^{\overline{q}} \varepsilon_t(C_q) u(\tilde{C} - C_q)}{\sum_{q=1}^{\overline{q}} u(\tilde{C} - C_q)} \xrightarrow[\overline{q}\to\infty]{P} 0,$$

where $u(\cdot)$ is the kernel smoothing function defined in (3.4), $\overline{q}$ is the total number of option contract prices observed at a sample time point t (i.e., $\overline{q}$ is the cross-sectional dimension of the panel of option prices) and $\tilde{C}$ denotes the maturity, money-ness pair of an option contract with characteristics $[\tilde{T}, \tilde{k}]$.

C1. Assumptions for Consistency

Given this structure of the observations, we impose the following assumptions to establish consistency of our estimator:

Assumption 1:

The stochastic process $\{X_t\}$ is a stationary Markov process of dimension $p \times 1$.

The distribution of $\left(X_{t_1}, X_{t_2}, \ldots, X_{t_n}, X_{t_{n+1}}\right)$ is indexed by a finite z–dimensional parameter vector $\theta \in \Theta \subset \mathbb{R}^z$, where Θ is compact.

Assumption 2 concerning our moment function $h_t(\tau;\theta)$ defined in (3.7):

$h_t(\tau;\theta)$ is a measurable function from $\mathbb{R}^{p\times 2} \times \mathbb{R}^p \times \mathbb{R}^p \times \Theta \to \mathbb{C}$.

$h_t(\tau;\theta)$ is a continuous function of θ on Θ.

[48] To differentiate between "noisy" implied volatility levels and the "true" ones, we use a superscript "0" in $\text{BSVol}^0\left(p_t(C)\right)$ to denote the true implied volatilities corresponding to option prices without microstructure noise.

[49] This zero-mean assumption should hold for each observed implied volatility quote, irrespective of the option contract's maturity and money-ness.

Assumption 3:

$\pi(\tau)$ is the probability density function of a distribution that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{p\times 2}$, satisfies $\pi(\tau) > 0$ for all $\tau \in \mathbb{R}^{p\times 2}$, and admits all moments.

We denote by $\mathbb{L}^2(\pi)$ the Hilbert space of complex valued functions that are square integrable with respect to π, that is:

$$\mathbb{L}^2(\pi) = \left\{ g : \mathbb{R}^{p\times 2} \to \mathbb{C} \mid \int |g(\tau)|^2 \pi(\tau)\, \mathrm{d}\tau < \infty \right\}.$$

The inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$ are defined on $\mathbb{L}^2(\pi)$ as follows:

$$\langle f, g \rangle \quad = \quad \int\limits_{\mathbb{R}^{p\times 2}} f(\tau)\overline{g(\tau)}\pi(\tau)\, \mathrm{d}\tau, \quad \text{and} \quad \|f\| = \langle f, f \rangle^{1/2}.$$

We assume that $h_t(\tau;\theta) \in \mathbb{L}^2(\pi), \forall \theta \in \Theta$ and $\forall \tau \in \mathbb{R}^{p\times 2}$.

Assumption 4:

The equation $\mathbb{E}^{\theta_0}\left[h_t(\tau;\theta)\right] = 0$, for $\forall \tau \in \mathbb{R}^{p\times 2}$ π-almost everywhere, has the unique solution $\theta_0 \in \mathrm{Int}(\Theta)$.

Here, $\mathbb{E}^{\theta_0}[\cdot]$ denotes the (unconditional) expectation with respect to the joint distribution of $[X_t; X_{t+1}]$ under the $\mathbb{P}$ measure characterized by θ_0.

Assumption 5:

$\|\hat{h}_n(\tau;\theta)\| \xrightarrow{P} \|\mathbb{E}^{\theta_0}\left[h_t(\tau;\theta)\right]\|$ uniformly on Θ.

Assumptions for Consistency of the 2nd Step Estimator

Assumption 6:

(i) $h_t(\tau;\theta)$ is continuously differentiable w.r.t. θ on Θ, $\forall \tau \in \mathbb{R}^{p\times 2}$.

(ii) $\mathbb{E}^{\theta_0}\left[\sup_{\theta\in\Theta} \|\nabla_\theta h_t(\cdot;\theta)\|\right] < \infty$.

Recall (3.11). Carrasco and Florens (2000) and Carrasco et al. (2007b) explain that the inverse

operator $K^{-1/2}$, given by $(K^{-1})^{1/2}$, does not exist on the whole space $\mathbb{L}^2(\pi)$, defined in Assumption 3 above, but only on a subset denoted by $\mathcal{H}(K)$ and given by the reproducing kernel Hilbert space associated with K. The inner product on $\mathcal{H}(K)$ is $\langle f, g\rangle_K = \langle K^{-1/2}f, K^{-1/2}g\rangle$, for $f, g \in \mathcal{H}(K)$. The following assumption is concerned with conditions for the existence of the reproducing kernel Hilbert space $\mathcal{H}(K)$ and the elements belonging to it.

Assumption 7:

(i) K, the asymptotic covariance operator of $\sqrt{n}\hat{h}_n(\theta_0)$, is a Hilbert-Schmidt operator.

(ii) The null space of K: $\mathcal{N}(K) = \{f \in \mathbb{L}^2(\pi) | Kf = 0\} = \{0\}$.

(iii) $\mathbb{E}^{\theta_0}\left[h_t(\tau;\theta)\right] \in \mathcal{H}(K), \forall\theta \in \Theta, \forall\tau \in \mathbb{R}^{p\times 2}$.

C2. Proof of Proposition 2

With the presupposed structure of our observations and given Assumptions 1 to 5, the proof of Proposition 2 follows from Theorem 3.3 in Gallant and White (1988).

To facilitate linking our assumptions to the assumptions required for the proof of Theorem 3.3. in Gallant and White (1988) (henceforth referred to as GW1988), we explicate the correspondence between Assumptions 1 to 5 and the assumptions required for Theorem 3.3 in GW1988 (employing the assumption names used in GW1988): Assumption 1 is contained in the GW1988 DGP assumption, while Assumptions 2 and 3 jointly ensure that the GW1988 optimand assumption is met. Assumption 4 corresponds to the GW1988 assumption regarding the identifiable uniqueness of the criterion minimizing parameter vector θ, while Assumption 5 corresponds to the GW1988 assumption for the uniform convergence of the criterion function on Θ. We discuss the validity of Assumptions 1 to 5 for the SVHJ model in Appendix C4.

C3. Proof of Proposition 3

The proof of Proposition 3 is similar to the proof of Proposition 2. However, two extra steps are required to establish the asymptotic properties of the (square root of the) regularized inverse estimator $(K_n^{\alpha_n})^{-1/2}$ of the covariance operator K, where α_n is a penalty that converges to 0

in probability as the sample size n grows to infinity.[50] These steps ensure that the regularized estimator converges in probability to a nearby operator of K which has a bounded inverse, an approximation for the theoretical covariance operator of the continuum of moment conditions evaluated at θ_0.

First, Assumptions 1 to 7 ensure that Proposition 3.3 (i) in Carrasco et al. (2007a) can be invoked to conclude that:

$$\|K_n - K\| \xrightarrow{P} 0.$$

Second, by Lemma B.2 provided in the appendix of Carrasco et al. (2007a), we can conclude that the Tikhonov regularized inverses of K and K_n, i.e., $(K^{\alpha_n})^{-1} = \left(K^2 + \alpha_n I\right)^{-1} K$ and $(K_n^{\alpha_n})^{-1} = \left(K_n^2 + \alpha_n I\right)^{-1} K_n$ satisfy the following convergence in probability:

$$\|(K_n^{\alpha_n})^{-1/2} - (K^{\alpha_n})^{-1/2}\| \xrightarrow{P} 0.$$

These two intermediary steps ensure that the use of the weighting operator $(K_n^{\alpha_n})^{-1/2}$ featured in the criterion function defined in equation (3.13) is admissible, see Lemma B.2(ii) in Carrasco et al. (2007a). Therefore, Assumptions 1 to 7 ensure that the conditions required for Theorem 3.3 in Gallant and White (1988) are met in the context of the criterion function with a weighting operator, and hence the weak consistency result for the second step estimator is obtained.

C4. Discussion on the Validity of the Assumptions for the SVHJ Model

This appendix presents some proofs and remarks which support the validity of the assumptions in Appendix C1 in the particular case of the SVHJ model.[51]

Assumption 1 - Under the SVHJ model, $\{X_t\}$ is an adapted stochastic process with dynamics defined by equations (2.4)–(2.6). It is stationary and Markov (under the relevant parameter restrictions discussed earlier). As the conditional characteristic function for the

[50] A (generic) discussion about the optimal choice of α_n for regularized inverses can be found in Carrasco and Kotchoni (2015).

[51] Due to the fact that the conditional characteristic function of the SVHJ model does not have a fully analytic expression, not all the assumptions can be given explicit proofs.

model under $\mathbb{P}$ is used in the estimation procedure, the moment conditions are a function of the full parameter set θ (i.e., the moment conditions also depend on the risk premium related parameters which appear in the $\mathbb{P}$-specification of the model). The transition density of this Markovian process under $\mathbb{P}$ is also a function of the full parameter set, implying that the distribution of $\left(X_{t_1}, X_{t_2}, \ldots, X_{t_n}, X_{t_{n+1}}\right)$ is indeed indexed by a finite z–dimensional parameter vector θ.

Assumption 2 - The properties of the $h_t(\tau;\cdot)$ moment function depend upon the properties of the state-implying function (inverse mapping) $X_{t_i}^{\theta} = f(P(X_{t_i}, \theta_0, C), \theta, C)$, as the latter serves as input in the moment function. The properties of this inverse mapping cannot be established analytically, as the pricing function is not available in closed form. The numerical implementation of the inverse mapping was well-behaved in all the simulations and empirical applications. The continuity of the moment function $h_t(\tau, \theta)$ in θ hinges on the continuity of the latent state-implying function $X_{t_i}^{\theta} = f(P(X_{t_i}, \theta_0, C), \theta, C)$ in $\theta^{\mathbb{Q}}$, which denotes the subset of parameters which appear in the $\mathbb{Q}$-specification of the model. Under additional mild assumptions which are required for the inverse function theorem to hold, such as the continuity of the partial derivatives of $f(P(X_{t_i}, \theta_0, C), \theta, C)$ with respect to the latent states, the state-implying function is a continuous function of the parameter vector $\theta \in \Theta$.

Assumption 3 - The following proof shows that the moment function is bounded and hence square integrable.
As the conditional characteristic function is bounded, i.e., $|\phi(s, X_{t_i}^{\theta}, \Delta; \theta)| \leq 1, \forall s \in \mathbb{R}^p$:

$$|h_t(\tau;\theta)| = |e^{\mathrm{i}r\cdot X_{t_i}^{\theta}}(e^{\mathrm{i}s\cdot X_{t_{i+1}}^{\theta}} - \phi(s, X_{t_i}^{\theta}, \Delta; \theta)| \leq |e^{\mathrm{i}(r\cdot X_{t_i}^{\theta} + s\cdot X_{t_{i+1}}^{\theta})}| + |e^{\mathrm{i}r\cdot X_{t_i}^{\theta}}\phi(s, X_{t_i}^{\theta}, \Delta; \theta)| \leq 2.$$

In our applications we choose $\pi(\cdot)$ to be the standard Gaussian multivariate density, which meets the conditions stated in Assumption 3. Therefore, $h_t(\tau;\theta) \in \mathbb{L}^2(\pi), \forall \theta \in \Theta$ and $\forall \tau \in \mathbb{R}^{p\times 2}$.

Assumption 4 - By virtue of the way in which the state-implying is defined, i.e., according to $X_{t_i}^{\theta_0} = f(P(X_{t_i}, \theta_0, C), \theta_0, C) = X_{t_i}$, the moment function evaluated at θ_0 becomes:

$$h_t(\tau;\theta_0) = e^{\mathrm{i}r\cdot X_{t_i}^{\theta_0}}(e^{\mathrm{i}s\cdot X_{t_{i+1}}^{\theta_0}} - \phi(s, X_{t_i}^{\theta_0}, \Delta; \theta_0)) = e^{\mathrm{i}r\cdot X_{t_i}}(e^{\mathrm{i}s\cdot X_{t_{i+1}}} - \phi(s, X_{t_i}, \Delta; \theta_0)).$$

Evaluated at the true parameter vector θ_0, the conditional characteristic function is:

$$\phi(s, X_{t_i}, \Delta; \theta_0) = \mathbb{E}^{\theta_0}\left[e^{\mathrm{i}s \cdot X_{t_{i+1}}} | X_{t_i}\right].$$

The expectation of the moment function evaluated at the true parameter vector θ_0 is:

$$\mathbb{E}^{\theta_0}\left[h_t(\tau;\theta)\right] = \mathbb{E}^{\theta_0}\left[e^{\mathrm{i}r \cdot X_{t_i}}\left(e^{\mathrm{i}s \cdot X_{t_{i+1}}} - \mathbb{E}^{\theta_0}\left[e^{\mathrm{i}s \cdot X_{t_{i+1}}} | X_{t_i}\right]\right)\right] = 0.$$

In the SVHJ model, X_{t_i} is a Markov process, hence using the properties of conditional expectations indeed

$$\mathbb{E}^{\theta_0}\left[h_t(\tau;\theta_0)\right] = 0, \ \forall \tau \in \mathbb{R}^{p\times 2}, \ \pi\text{-almost everywhere.}$$

Providing an analytic proof for the solution's uniqueness at the interior point $\theta_0 \in \mathrm{Int}(\Theta)$ is beyond the scope of this discussion given the absence of an analytic expression for the state-implying function.

Assumption 5 - In order to prove the uniform convergence of the criterion function, we invoke an intermediate result from GW1988, Lemma 3.4, according to which, given the set-up of our Hilbert space given in Assumption 3, it is enough to prove that the moment function converges (uniformly in $\theta \in \Theta$). To prove the latter we refer to a functional central limit theorem from Politis and Romano (1994). This limit theorem by Politis and Romano is reproduced here with a slight change in notation to match our set-up:

Theorem 2.2 (in Politis and Romano (1994)) Assume that the $h_{t_i}(\tau;\theta)$ are stationary, bounded with probability one, and have α-mixing coefficients such that $\sum_{i=1}^{j} i^2\alpha(i) \leq Kj^r$, for all $1 \leq j \leq t_n$ where $r < 3/2$ and K is a positive constant.[52] Then, $\sqrt{n}\hat{h}_n(\tau;\theta_0)$ converges weakly to a Gaussian random element of $\mathbb{L}^2(\pi)$ with a zero mean.

Given that the $h_{t_i}(\tau;\theta)$ are stationary (by Assumption 1 and Assumption 2) and bounded as proven, the only necessary prerequisite for the application of this central limit theorem is an investigation of the mixing properties of the moment function. According to Assumption 2, the moment function is measurable, so it inherits the mixing properties of the state variables.

[52] $\alpha(j) = \sup_{A,B}|\mathbb{P}[AB] - \mathbb{P}[A]\mathbb{P}[B]|$, where A, B vary in the σ-algebras generated by $\{h_{t_i}, t_i \leq k\}$ and $\{h_{t_i}, t_i \geq j+k\}$, respectively, for any $k \geq 1$.

Following Liebscher (2005), we argue that as X_t is a stationary Markov process, under the additional assumption that it is also geometrically ergodic, the α-mixing property required in the theorem by Politis and Romano is fulfilled. To establish the geometric ergodicity property for X_t, one could resort to a geometric drift condition as in Meyn and Tweedie (2009), p. 367.

Uniform convergence in θ can then be established by invoking Lemma B.3 from Carrasco et al. (2007a). Explicitly proving the stochastic equicontinuity required as a pre-condition in that lemma would require knowledge of the score of the state-implying function and the score of the conditional characteristic function, neither of which are available in fully analytic form for the SVHJ model.

Assumption 6 - The differentiability conditions are required to establish the convergence of the estimator of the covariance operator to its theoretical counterpart. The continuous differentiability of the moment function $h_t(\tau;\theta)$ in θ is inherited from the continuous differentiability of the state-implying function in $\theta^{\mathbb{Q}}$ and from the continuous differentiability of the conditional characteristic function in θ.

Assumption 7 - A sufficient condition such that the operator K is a Hilbert-Schmidt operator is to have that $\int\int |k(\tau_1,\tau_2)|\pi(\tau_1)\pi(\tau_2)\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2 < \infty$. As mentioned in Politis and Romano (1994), this holds if $h_t(\tau;\theta)$ is bounded $\forall\tau \in \mathbb{R}^{p\times 2}$ and $\forall\theta \in \Theta$, and if $h_t(\tau;\theta)$ has $\alpha(\cdot)$-mixing coefficients which are summable. Both conditions have been discussed above.

Appendix D

Options Data-Set: Sample Span and Descriptive Statistics

[Table 3]

[Table 4]

Appendix E

SVJ and SVVJ Model Specifications and Parameter Estimates

Employing the same notation used to describe the SVHJ model in equations (2.4)–(2.6) and (2.11)–(2.13), the stochastic volatility with Poissonian jump component model, labeled SVJ,

and the stochastic volatility with volatility driven jump intensity model, labeled SVVJ, are specified below under the equivalent probability measures $\mathbb{P}$ and $\mathbb{Q}$. The model specific measure changes are generated by candidate pricing kernels similar to the one defined in Appendix B.

SVJ

The SVJ model specification under $\mathbb{P}$ is:

$$\begin{aligned} dy_t &= \left(\left(\eta - \frac{1}{2}\right) v_t + (\mu - \mu^*)\lambda\right) \mathrm{d}t + \sqrt{v_t}\, \mathrm{d}W_t^{(1),\mathbb{P}} + \mathrm{d}J_t^{\mathbb{P}} - \mu\lambda\, \mathrm{d}t; \\ dv_t &= \kappa_v \left(\overline{v} - v_t\right) \mathrm{d}t + \sigma_v \sqrt{v_t} \left(\rho\, \mathrm{d}W_t^{(1),\mathbb{P}} + \sqrt{1-\rho^2}\, \mathrm{d}W_t^{(2),\mathbb{P}}\right); \end{aligned}$$

where $\mathrm{d}J_t^{\mathbb{P}} = Z_t^{\mathbb{P}}\, \mathrm{d}N_t$, with $Z_t^{\mathbb{P}} \sim \mathcal{N}\left(\mu_j^{\mathbb{P}}, \sigma_j^2\right)$. The SVJ model specification under $\mathbb{Q}$ is:

$$\begin{aligned} dy_t &= -\frac{1}{2} v_t\, \mathrm{d}t + \sqrt{v_t}\, \mathrm{d}W_t^{(1),\mathbb{Q}} + \mathrm{d}J_t^{\mathbb{Q}} - \mu^*\lambda\, \mathrm{d}t; \\ dv_t &= \kappa_v \left(\overline{v} - v_t\right) \mathrm{d}t + \sigma_v \sqrt{v_t} \left(\rho\, \mathrm{d}W_t^{(1),\mathbb{Q}} + \sqrt{1-\rho^2}\, \mathrm{d}W_t^{(2),\mathbb{Q}}\right); \end{aligned}$$

where $\mathrm{d}J_t^{\mathbb{Q}} = Z_t^{\mathbb{Q}}\, \mathrm{d}N_t$, with $Z_t^{\mathbb{Q}} \sim \mathcal{N}\left(\mu_j^{\mathbb{Q}}, \sigma_j^2\right)$. Here, N_t denotes a standard Poisson counting process with constant intensity $\lambda \in \mathbb{R}^+$ under both $\mathbb{P}$ and $\mathbb{Q}$.

SVVJ

The SVVJ model specification under $\mathbb{P}$ is:

$$\begin{aligned} dy_t &= \left(\left(\eta - \frac{1}{2}\right) v_t + (\mu - \mu^*)\lambda_1 v_t\right) \mathrm{d}t + \sqrt{v_t}\, \mathrm{d}W_t^{(1),\mathbb{P}} + \mathrm{d}J_t^{\mathbb{P}} - \mu\lambda_1 v_t\, \mathrm{d}t; \\ dv_t &= \kappa_v \left(\overline{v} - v_t\right) \mathrm{d}t + \sigma_v \sqrt{v_t} \left(\rho\, \mathrm{d}W_t^{(1),\mathbb{P}} + \sqrt{1-\rho^2}\, \mathrm{d}W_t^{(2),\mathbb{P}}\right); \end{aligned}$$

where $\mathrm{d}J_t^{\mathbb{P}} = Z_t^{\mathbb{P}}\, \mathrm{d}N_t$, with $Z_t^{\mathbb{P}} \sim \mathcal{N}\left(\mu_j^{\mathbb{P}}, \sigma_j^2\right)$. The SVVJ model specification under $\mathbb{Q}$ is:

$$\begin{aligned} dy_t &= -\frac{1}{2} v_t\, \mathrm{d}t + \sqrt{v_t}\, \mathrm{d}W_t^{(1),\mathbb{Q}} + \mathrm{d}J_t^{\mathbb{Q}} - \mu^*\lambda_1 v_t\, \mathrm{d}t \\ dv_t &= \kappa_v \left(\overline{v} - v_t\right) \mathrm{d}t + \sigma_v \sqrt{v_t} \left(\rho\, \mathrm{d}W_t^{(1),\mathbb{Q}} + \sqrt{1-\rho^2}\, \mathrm{d}W_t^{(2),\mathbb{Q}}\right); \end{aligned}$$

where $\mathrm{d}J_t^{\mathbb{Q}} = Z_t^{\mathbb{Q}}\,\mathrm{d}N_t$, with $Z_t^{\mathbb{Q}} \sim \mathcal{N}\left(\mu_j^{\mathbb{Q}}, \sigma_j^2\right)$. Here, N_t denotes a counting process with time-varying jump intensity $\lambda_t = \lambda_1 v_t, \lambda_1 \in \mathbb{R}^+$ under both $\mathbb{P}$ and $\mathbb{Q}$.

The parameter estimates for the SVJ and SVVJ specifications are:

[Table 5]

Appendix F

Equity Risk Premium (ERP)

The equity risk premium for the time span $[t, T]$ is defined as:

$$\mathrm{ERP}_t = \frac{1}{T-t}\left(\mathbb{E}_t\left[\frac{F_T - F_t}{F_t}\right] - \mathbb{E}_t^{\mathbb{Q}}\left[\frac{F_T - F_t}{F_t}\right]\right).$$

Because the futures price is a martingale under the risk neutral measure, the last term in the expression for ERP_t is zero, i.e.:

$$\begin{aligned}\mathrm{ERP}_t &= \frac{1}{T-t}\left(\mathbb{E}_t\left[\frac{F_T - F_t}{F_t}\right] - \underbrace{\mathbb{E}_t^{\mathbb{Q}}\left[\frac{F_T - F_t}{F_t}\right]}_{=0}\right) = \frac{1}{T-t}\mathbb{E}_t\left[\frac{F_T - F_t}{F_t}\right] \\ &= \frac{1}{T-t}\mathbb{E}_t\Big[\int_t^T \left(\eta v_s + (\mu - \mu^*)\lambda_s\right)\mathrm{d}s\Big].\end{aligned}$$

By plugging in the analytic expressions for the $\mathcal{F}_t$-conditional expectations of v_s and λ_s, $s \geq t$, and calculating the integrals, we derive the total equity risk premium in our model:

$$\begin{aligned}
\mathrm{ERP}_t &= \frac{1}{T-t}\mathbb{E}_t\left[\int_t^T \eta v_s \,\mathrm{d}s\right] + \frac{1}{T-t}\mathbb{E}_t\left[\int_t^T (\mu-\mu^*)\lambda_s \,\mathrm{d}s\right] \\
&= \frac{1}{T-t}\int_t^T \eta\mathbb{E}_t\left[v_s\right]\mathrm{d}s + \frac{1}{T-t}\int_t^T (\mu-\mu^*)\mathbb{E}_t\left[\lambda_s\right]\mathrm{d}s \\
&= \frac{1}{T-t}\int_t^T \eta\left(\overline{v} + e^{-\kappa_v(s-t)}\left(v_t-\overline{v}\right)\right)\mathrm{d}s \\
&\qquad + \frac{1}{T-t}\int_t^T (\mu-\mu^*)\left(\frac{\kappa_\lambda\overline{\lambda}}{\kappa_\lambda-\delta}\left(1-e^{-(\kappa_\lambda-\delta)(s-t)}\right) + e^{-(\kappa_\lambda-\delta)(s-t)}\lambda_t\right)\mathrm{d}s \\
&= \eta\left(\overline{v} + \frac{v_t-\overline{v}}{\kappa_v(T-t)}\left(1-e^{-\kappa_v(T-t)}\right)\right) \\
&\quad + \left(\mu-\mu^*\right)\left(\frac{\kappa_\lambda\overline{\lambda}}{\kappa_\lambda-\delta} + \left(1-e^{-(\kappa_\lambda-\delta)(T-t)}\right)\frac{\left(\kappa_\lambda-\delta\right)\lambda_t - \kappa_\lambda\overline{\lambda}}{\left(\kappa_\lambda-\delta\right)^2(T-t)}\right).
\end{aligned}$$

The analytic expressions for the conditional moments of the state variables were derived using the martingale property of the infinitesimal generator.

Variance Risk Premium (VRP)

The variance risk premium for the time span $[t,T]$ is given by:

$$\begin{aligned}
\mathrm{VRP}_t &= \frac{1}{T-t}\left(\mathbb{E}_t\left[\mathrm{QV}_{[t,T]}\right] - \mathbb{E}_t^{\mathbb{Q}}\left[\mathrm{QV}_{[t,T]}\right]\right) \\
&= \frac{1}{T-t}\left(\mathbb{E}_t\left[\int_t^T v_s \,\mathrm{d}s + \int_t^T Z_s^2 \,\mathrm{d}N_s\right] - \mathbb{E}_t^{\mathbb{Q}}\left[\int_t^T v_s \,\mathrm{d}s + \int_t^T Z_s^2 \,\mathrm{d}N_s\right]\right) \\
&= \frac{1}{T-t}\left(\mathbb{E}\left[Z_t^2\right] - \mathbb{E}^{\mathbb{Q}}\left[Z_t^2\right]\right)\int_t^T \mathbb{E}_t\left[\lambda_s\right]\mathrm{d}s \\
&= \left(\left(\mu_j^{\mathbb{P}}\right)^2 - \left(\mu_j^{\mathbb{Q}}\right)^2\right)\left(\frac{\kappa_\lambda\overline{\lambda}}{\kappa_\lambda-\delta} + \left(1-e^{-(\kappa_\lambda-\delta)(T-t)}\right)\frac{\left(\kappa_\lambda-\delta\right)\lambda_t - \kappa_\lambda\overline{\lambda}}{\left(\kappa_\lambda-\delta\right)^2(T-t)}\right),
\end{aligned}$$

exploiting the analytic expressions for the conditional moments of the state variables derived above.

Appendix G

Jump Risk Premium: Option Pricing Impact

[Table 6]

References

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. A. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606.

Aït-Sahalia, Y., Laeven, R. J. A., and Pelizzon, L. (2014). Mutual excitation in Eurozone sovereign CDS. *Journal of Econometrics*, 183(2):151–167.

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53(2):499–547.

Andersen, T. G., Benzoni, L., and Lund, J. (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance*, 57(3):1239–1284.

Andersen, T. G., Fusari, N., and Todorov, V. (2015a). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

Andersen, T. G., Fusari, N., and Todorov, V. (2015b). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558–584.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5):2003–2049.

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94(1-2):181–238.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section*. American Statistical Association.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Bollerslev, T. and Todorov, V. (2011). Tails, fears, and risk premia. *Journal of Finance*, 66(6):2165–2211.

Bollerslev, T., Todorov, V., and Xu, L. (2015). Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1):113–134.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141:876–912.

Broadie, M., Chernov, M., and Johannes, M. (2007). Model specification and risk premia: evidence from futures options. *Journal of Finance*, 62(3):1453–1490.

Broadie, M., Chernov, M., and Johannes, M. (2009). Understanding index option returns. *Review of Financial Studies*, 22(11):4493–4529.

Carr, P. and Madan, D. (1999). Option valuation using the Fast Fourier Transform. *Journal of Computational Finance*, 2(4):61–73.

Carrasco, M., Chernov, M., Florens, J.-P., and Ghysels, E. (2007a). Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140(2):529–573.

Carrasco, M. and Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, 16(06):797–834.

Carrasco, M. and Florens, J.-P. (2002). Efficient GMM estimation using the empirical characteristic function. *IDEI Working paper*, 140.

Carrasco, M., Florens, J.-P., and Renault, E. (2007b). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of Econometrics*, 6:5633–5751.

Carrasco, M. and Kotchoni, R. (2015). Efficient estimation using the characteristic function. *Working paper*.

CBOE (2009). The CBOE volatility index – VIX. White Paper.

Cont, R. and Da Fonseca, J. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Duffie, D., Pan, J., and Singleton, K. J. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Duffie, D. and Singleton, K. J. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica*, 61(4):929–952.

Egloff, D., Leippold, M., and Wu, L. (2010). Valuation and optimal investing in variance swaps. *Journal of Financial and Quantitative Analysis*, 45(5):1279–1310.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance*, 59(3):1367–1404.

Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, 58(3):1269–1300.

Errais, E., Giesecke, K., and Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665.

Fan, Y., Pastorello, S., and Renault, E. (2015). Maximization by parts in extremum estimation. *The Econometrics Journal*, 18(2):147–171.

Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182.

Gallant, A. R. and White, H. (1988). *A unified theory of estimation and inference for nonlinear dynamic models*. Basil Blackwell New York.

Hansen, L. P. (1985). A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics*, 30(1):203–238.

Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(3):438–443.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Johannes, M. S., Polson, N. G., and Stroud, J. R. (2009). Optimal filtering of jump diffusions: extracting latent states from asset prices. *Review of Financial Studies*, 22(7):2759–2799.

Li, G. and Zhang, C. (2013). Conditional jump intensity, conditional expected jump size, and relative stock price level. *Working paper*.

Liebscher, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis*, 26(5):669–689.

Lord, R., Koekkoek, R., and Dijk, D. V. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194.

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125–144.

Meyn, S. P. and Tweedie, R. L. (2009). *Markov chains and stochastic stability*. Cambridge University Press.

Pan, J. (2002). The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):3–50.

Pastorello, S., Patilea, V., and Renault, E. (2003). Iterative and recursive estimation in structural nonadaptive models. *Journal of Business & Economic Statistics*, 21(4):449–509.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Protter, P. E. (2005). *Stochastic integration and differential equations*. Springer.

Schneider, P. G. and Trojani, F. (2015). Fear trading. *Working paper*.

Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102(1):111–141.

Taylor, R. L. (1978). *Stochastic convergence of weighted sums of random elements in linear spaces*. Springer.

Whaley, R. E. (2009). Understanding the VIX. *The Journal of Portfolio Management*, 35(3):98–105.

Tables

Table 1: Monte Carlo simulation results based on 100 simulated option panels each containing 500 observations (frequency set to match weekly data). True parameter values, Monte Carlo sample means, standard deviations, and 10%, 25%, 50%, 75%, and 90% quantiles of the empirical distribution of each model parameter are presented on separate rows.

	$\mu_J^{\mathbb{P}}$	$\mu_J^{\mathbb{Q}}$	σ_J	η	κ_v	$\overline{v}$	σ_v	ρ	κ_λ	$\bar{\lambda}$	δ
True	-5%	-14%	6%	2.40	4.80	0.01	0.22	-0.60	18.00	0.30	16.5
MC Mean	-7.51%	-14.13%	5.31%	2.87	5.31	0.01	0.23	-0.61	22.51	0.33	16.26
MC S.D.	3.03%	1.57%	1.96%	0.83	0.92	0.001	0.03	0.08	4.46	0.03	2.65
10%	-11.16%	-16.33%	3.37%	1.86	4.30	0.009	0.19	-0.72	16.52	0.28	13.43
25%	-9.48%	-14.91%	4.88%	2.25	4.71	0.009	0.21	-0.68	18.87	0.31	14.64
50%	-7.87%	-14.12%	5.90%	2.88	5.18	0.01	0.23	-0.62	22.84	0.33	16.17
75%	-5.34%	-13.16%	7.34%	3.46	6.16	0.012	0.24	-0.54	25.94	0.35	17.88
90%	-3.73%	-11.94%	8.15%	3.97	6.61	0.013	0.26	-0.51	28.06	0.38	19.54

Table 2: Estimated model parameters and standard errors (in parentheses). $(*), (**)$ and $(***)$ denote statistical significance at the conventional $10\%, 5\%$ and 1% significance levels.

$\mu_J^{\mathbb{P}}$	$\mu_J^{\mathbb{Q}}$	σ_J	η	κ_v	$\overline{v}$	σ_v	ρ	κ_λ	$\bar{\lambda}$	δ
$-4.86\%^{***}$	$-13.68\%^{***}$	$6.63\%^{**}$	2.37^{*}	4.76^{***}	0.011^{***}	0.225^{***}	-0.61^{***}	18.16^{***}	0.326^{***}	16.62^{***}
(1.48%)	(1.69%)	(2.94%)	(1.4)	(1.25)	(0.002)	(0.08)	(0.24)	(1.65)	(0.017)	(2.51)

Table 3: Total, in-sample, out-of-sample spans

Sample:	No. of days:	Start date:	End date:	No. of options:
Total sample	913 Wednesdays	1 January 1996	31 August 2013	259691 options
In-sample	723 Wednesdays	1 January 1996	31 December 2009	145790 options
Out-of-sample	190 Wednesdays	1 January 2010	31 August 2013	113901 options

Table 4: Descriptive statistics for the in- and out-of-sample options. The option set contains the actual at- and out-of-the-money call option prices recorded and the "synthetic" in-the-money call option prices generated using the at-the-money put-call parity with the corresponding maturity forward price as underlying.

In-sample options						Percentiles			
Variable	Mean	Std. dev.	Min	5%	25%	50%	75%	95%	Max
Call Price ($)	128.25	138.38	0.12	0.85	20.01	77.66	196.34	417.00	774.43
Implied Vol.	0.2376	0.0791	0.0814	0.1436	0.1800	0.2202	0.2773	0.3905	0.8618
Maturity	0.46	0.31	0.09	0.09	0.19	0.36	0.71	1.04	1.18
Money-ness (K/S)	0.96	0.18	0.50	0.64	0.84	0.97	1.07	1.27	1.49
Out-of-sample options						Percentiles			
Variable	Mean	Std. dev.	Min	5%	25%	50%	75%	95%	Max
Call Price ($)	138.63	146.17	0.125	0.55	15.75	85.91	224.02	444.19	645.61
Implied Vol.	0.1715	0.0721	0.0755	0.0958	0.1180	0.1520	0.2059	0.3038	0.7955
Maturity	0.44	0.31	0.09	0.09	0.17	0.34	0.72	1.02	1.18
Money-ness (K/S)	0.92	0.15	0.50	0.62	0.82	0.94	1.03	1.16	1.29

Table 5: SVJ and SVVJ parameter estimates

SVJ								
$\mu_j^{\mathbb{P}}$	$\mu_j^{\mathbb{Q}}$	σ_j	η	κ_v	$\overline{v}$	σ_v	ρ	λ
-13.21%	-18.77%	2.62%	2.47	5.13	0.009	0.195	-0.40	1.14
SVVJ								
$\mu_j^{\mathbb{P}}$	$\mu_j^{\mathbb{Q}}$	σ_j	η	κ_v	$\overline{v}$	σ_v	ρ	λ_1
-3.87%	-21.70%	3.70%	2.89	4.22	0.014	0.345	-0.45	28.12

Table 6: Relative pricing impact of the jump risk premium for different combinations of money-ness levels and maturities. The jump risk premium increases return volatility. Therefore, most call options become more expensive, except for the out-of-the-money contracts. The latter (bottom row) become cheaper due to the increase in the skewness of the return distribution induced by the jump risk premium.

	Maturity			
Money-ness	0.25	0.5	0.75	1
95%	1.4%	1.6%	2.0%	2.9%
97.5%	2.0%	2.1%	2.5%	3.5%
100 %	3.1%	3.3%	3.7%	4.6%
102.5%	4.0%	4.2%	4.6%	5.5%
105%	2.3%	2.5%	2.9%	3.8%
107.5 %	-13.1%	-13.0%	-12.6%	-11.6%

Figures

Figure 1: Time series of option-implied stochastic volatility and stochastic jump intensity. The upper panel plots the option-implied annualized volatility process, $\sqrt{v_t}$, expressed in % points and the log-returns on the S&P 500 index (secondary, right-hand axis). The lower panel plots the option-implied jump intensity process, λ_t, and the level of the S&P 500 index (secondary, right-hand axis). The out-of-sample-period latent state levels are implied from out-of-sample option data using parameter estimates obtained from in-sample option data.

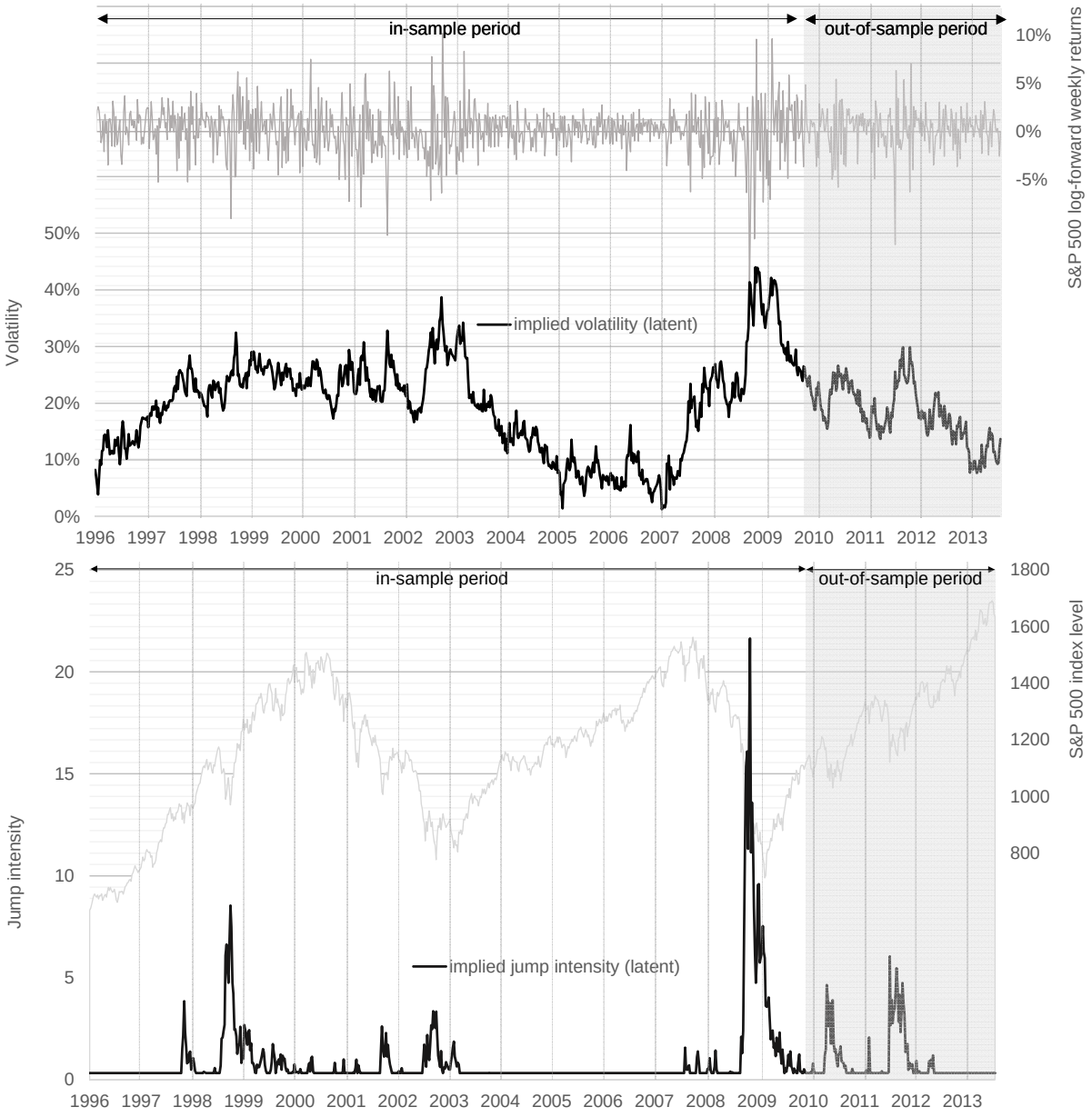


Figure 2: Implied volatility surface dynamics after the occurrence of a self-exciting jump. This figure depicts the time decay of the impact of a jump on option prices, by plotting the consecutive implied volatility surfaces at 1-month time intervals after an initial jump at time $t = 0$, assuming that no further jumps occur after $t = 0$. The latent volatility state is set to its long run mean level, i.e., $v_0 \equiv \cdots \equiv v_{4\mathrm{M}} \equiv \overline{v}$.

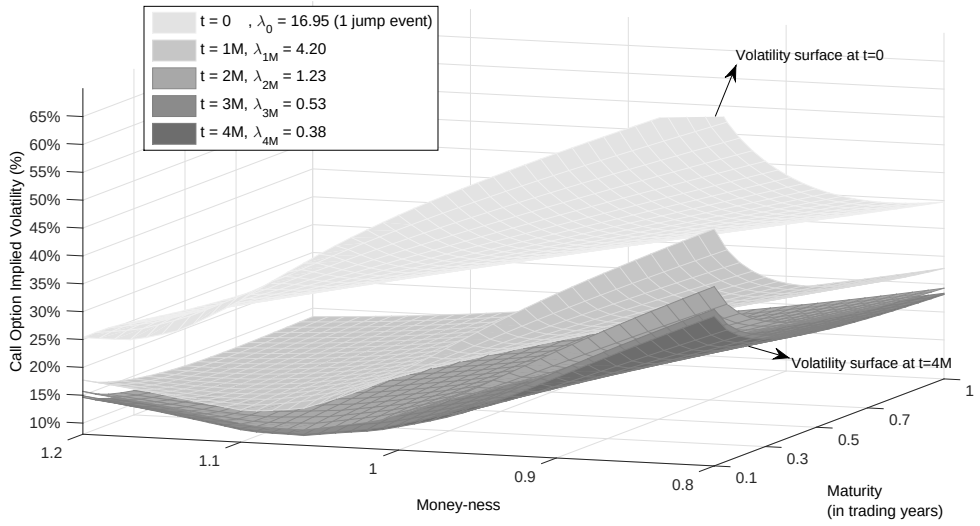


Figure 3: Mean Pricing Errors. Each quadrant plots the mean pricing errors (in implied volatility % points) for each of the three models (SVHJ, SVVJ, and SVJ). Pricing errors are computed for each model by averaging the option pricing errors over all sample days, for the option contracts with the maturity and money-ness characteristics displayed on the horizontal and vertical axes.

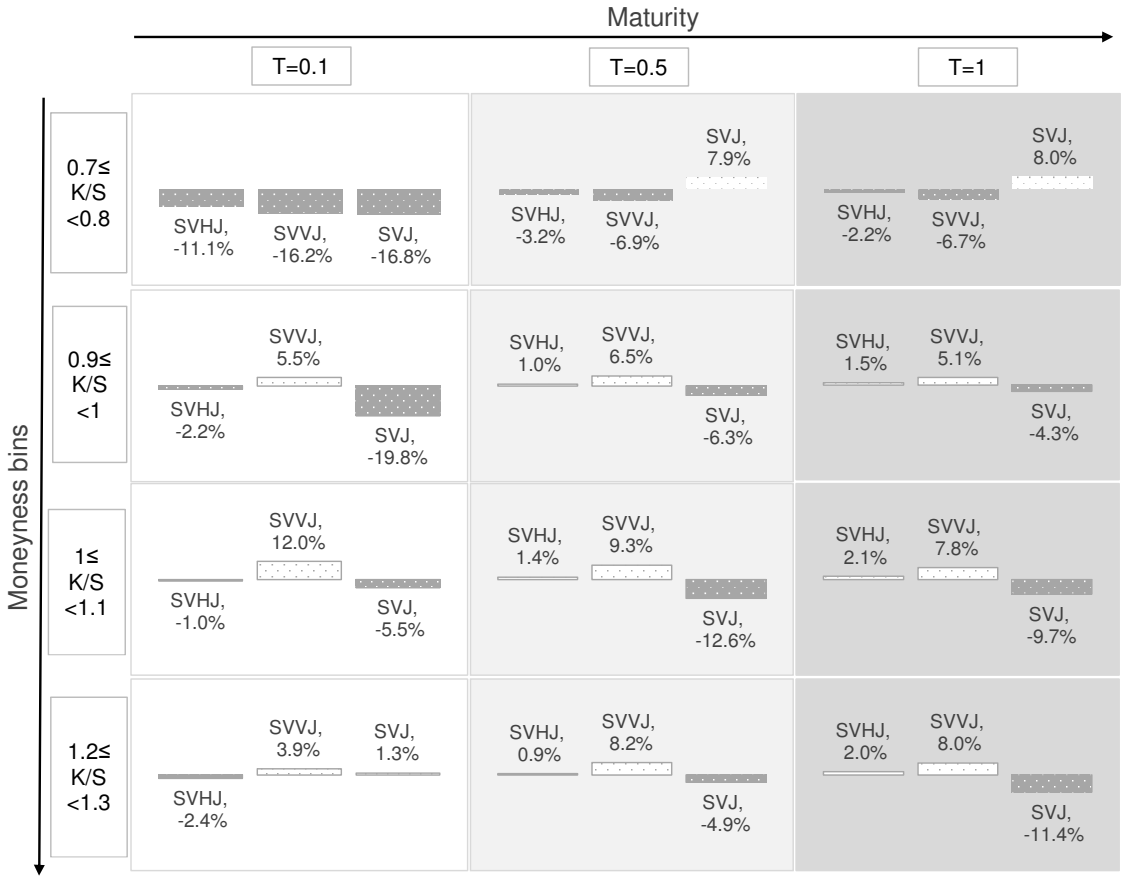


Figure 4: Root Mean Squared Pricing Errors. Each quadrant plots the root mean squared pricing errors (in implied volatility % points) for each of the three models (SVHJ, SVVJ, and SVJ). Pricing errors are computed for each model by averaging the squared option pricing errors over all sample days, for the option contracts with the maturity and money-ness characteristics displayed on the horizontal and vertical axes.

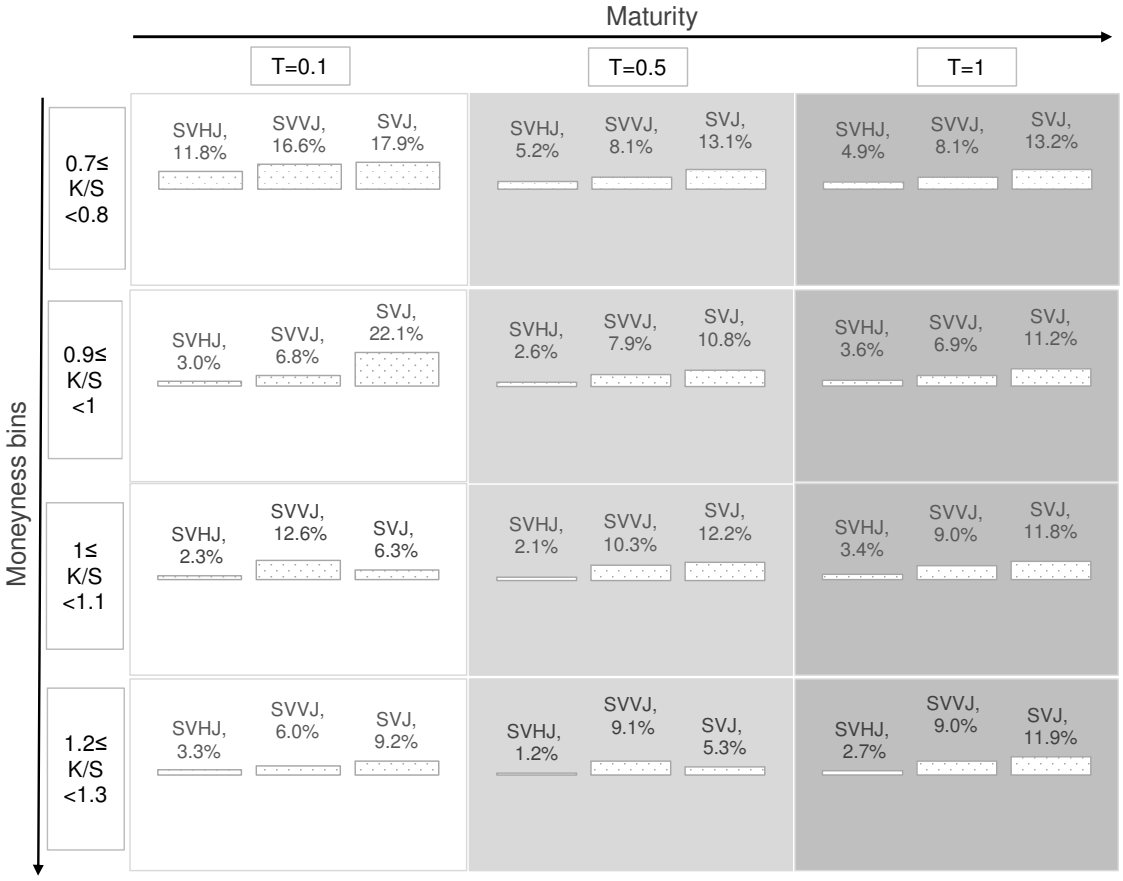


Figure 5: Equity risk premiums: $\mathrm{ERP}_t, \mathrm{ERP}_t^{\mathrm{dif}}$ and $\mathrm{ERP}_t^{\mathrm{jump}}$. The upper panel (secondary, right-hand axis) illustrates the 1-month ($T = t + 0.08$) jump risk premium ($\mathrm{ERP}_t^{\mathrm{jump}}$), while the lower panel shows the 1-month equity risk premium (ERP_t, thick black line) and the 1-month diffusive risk premium ($\mathrm{ERP}_t^{\mathrm{dif}}$, grey shaded area), all in % per annum. The out-of-sample period risk premiums are computed using latent state levels which are implied from out-of-sample option data using parameter estimates obtained from in-sample option data.

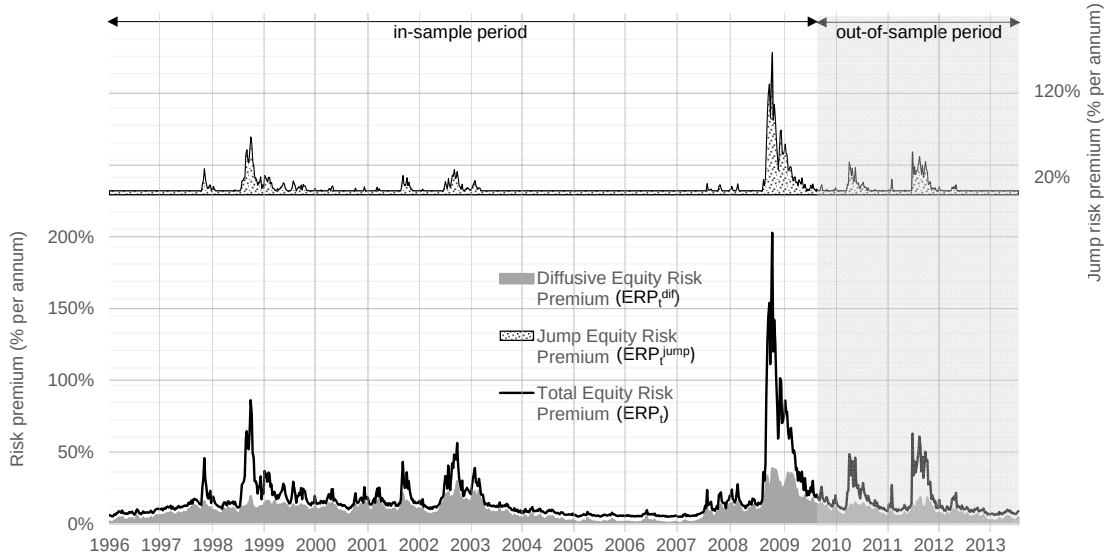


Figure 6: Implied volatility surfaces based on option prices from the SVHJ model. This figure plots implied volatility surfaces based on option prices computed from the SVHJ model, with compensation for jump size risk and with compensation for jump size risk "turned off" by replacing $\mu_j^{\mathbb{Q}}$ by $\mu_j^{\mathbb{P}}$. The latent states, λ_t and v_t, are set equal to their long run means.

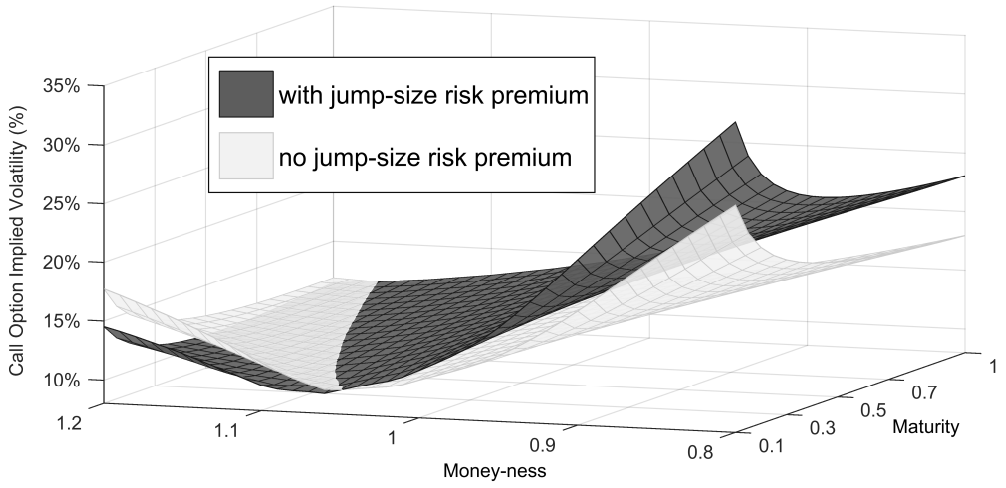


Figure 7: Investor fear index based on the variance risk premium. This figure plots the model-implied variance risk premium calculated for every week of the data sample from 1996 to 2013 along with the S&P 500 price index level. The expectations implicit in the variance risk premium are computed over the next month. The out-of-sample period variance risk premium is computed using latent state levels which are implied from out-of-sample option data using parameter estimates obtained from in-sample option data. The figure shows a smoothed version of the model-implied variance risk premium (VRP), i.e., a moving average over four observations (roughly one month) to allow for a comparison with the results in Bollerslev and Todorov (2011) that are included in the figure (dotted line).

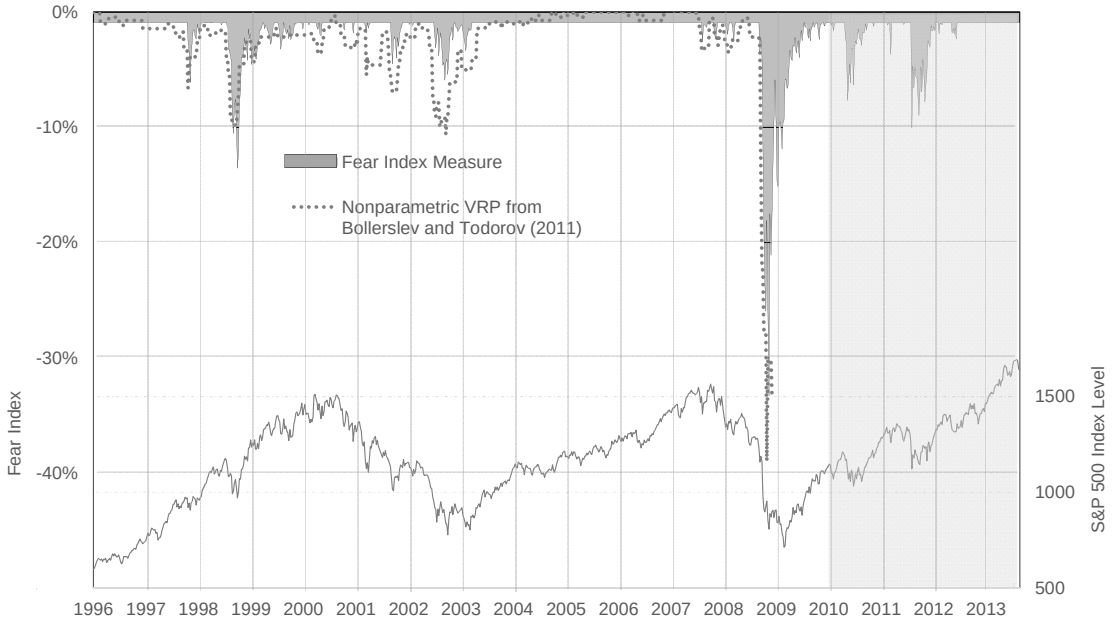
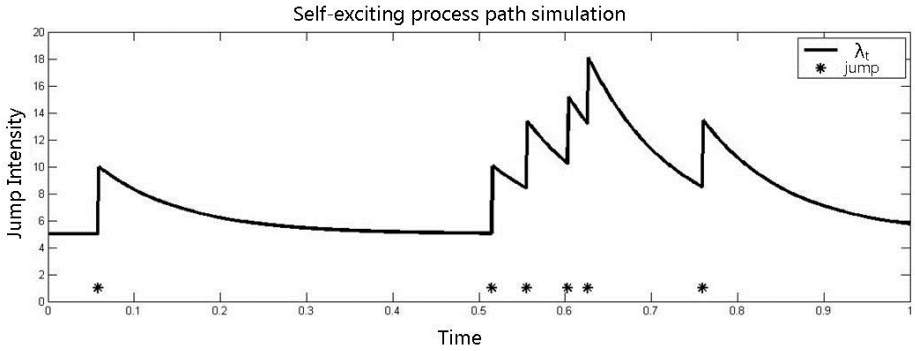
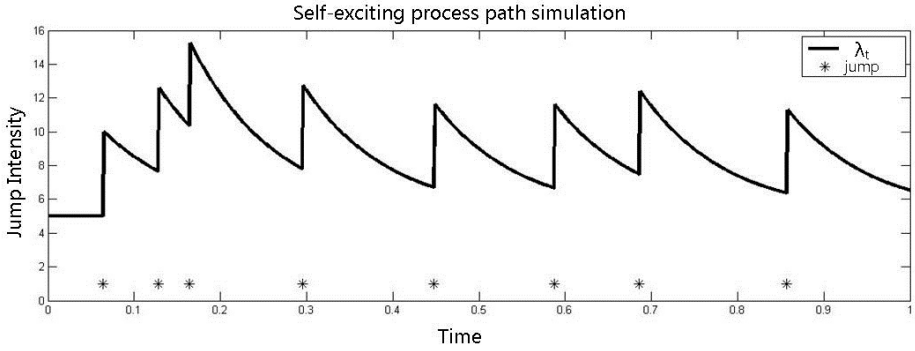


Figure 8: Simulated sample paths for the jump intensity process, λ_t, and the counter, N_t, of a self-exciting jump process with the specified parameters and time horizon.

(a) $\bar{\lambda} = 5;\ \kappa_\lambda = 10;\ \delta = 5;\ T = 1.$



(b) $\bar{\lambda} = 5;\ \kappa_\lambda = 2;\ \delta = 1;\ T = 1.$



(c) $\bar{\lambda} = 5;\ \kappa_\lambda = 15;\ \delta = 10;\ T = 10.$