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

H. Peter Boswijk†  Roger J. A. Laeven‡  Andrei Lalu§

January 18, 2015

Abstract

We propose an option pricing model with a self-exciting jump component inducing jump clustering, a phenomenon that is empirically relevant when financial markets are in turmoil. We develop a procedure to filter out the latent state variables of the model from a panel of option prices and estimate the model parameters via the generalized method of moments. We employ a continuum of moment conditions derived from the model’s conditional characteristic function and prove that, under natural assumptions, this estimation approach is consistent. Monte Carlo simulations show that our estimation procedure has good performance in finite samples. Based on a long time-series of S&P 500 options prices, we find strong evidence of self-excitation. The in-sample and out-of-sample option pricing performance of our model with self-exciting jumps is superior to that of alternative models with (possibly) time-varying jump intensity specifications.

Keywords: Hawkes processes, option pricing, affine jump diffusions, conditional characteristic function, GMM, continuum of moments.

JEL Codes: G12; C32; C58.

---

*This research was funded in part by the Netherlands Organization for Scientific Research (NWO) under grant NWO Vidi 2009 (Laeven).
†University of Amsterdam, Tinbergen Institute, Valckenierstraat 65, 1018 XE Amsterdam, Netherlands. E-mail: H.P.Boswijk@uva.nl.
‡University of Amsterdam and EURANDOM, Valckenierstraat 65, 1018 XE Amsterdam, Netherlands. E-mail: R.J.A.Laeven@uva.nl.
§University of Amsterdam, Tinbergen Institute and Duisenberg School of Finance, Valckenierstraat 65, 1018 XE Amsterdam, Netherlands. E-mail: A.Lalu@uva.nl.
1 Introduction

Existing research in financial econometrics documents the better fit achieved by using stochastic volatility and jumps to model asset returns and option price surfaces in continuous-time. For example, compared to a standard continuous Brownian diffusion, the prototypical Poissonian jump-diffusion that dates back to Merton (1976) yields a better fit for short time-to-maturity option prices. Yet, when financial markets are in turmoil, asset price crashes occur more frequently than predicted by models with standard stochastic volatility and jump components. Furthermore, the model-implied option price surfaces 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 replicate this pattern of crash clustering. Therefore, alternative model specifications are required to better accommodate the empirical patterns found in asset returns and lead to an improved fit for option pricing models.

Recently, Aït-Sahalia et al. (2014a) introduced a novel way of modeling the complex interplay between jumps over time and across different asset markets by introducing mutually exciting jump processes, also known as Hawkes processes after Hawkes (1971), into a semimartingale asset return model with a drift and stochastic volatility component. Aït-Sahalia et al. (2014a) find evidence 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., Bacry et al., 2013). Aït-Sahalia et al. (2014b) 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 threefold. First, we extend existing 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 and testing
procedure for this model, which first involves backing out the models 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. 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.

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-Scholes 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). 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.

To build our continuum of moment conditions, we use 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 first proposed by Duffie, Pan & Singleton (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 the likelihood function, and by choosing an appropriate set of instruments, consistency of the estimators may be ensured. 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). 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 our approach allows for the presence of latent variables: stochastic volatility and stochastic jump intensity. We show that, under natural assumptions, our estimation procedure is consistent even when the state vector contains latent states that are implied from option prices. 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.

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\footnote{Other risk factors such as interest rate uncertainty, dividend uncertainty or liquidity concerns could potentially have a very different impact on option prices. We focus on jump risks and diffusive risks to investigate the differences and the dynamic behavior of the risk premiums originating from these two risk sources.}, the estimation procedure we develop provides a way to achieve the desirable identification and efficiency features of GMM estimation with a continuum of moment conditions while simultaneously estimating the model under both the physical and the risk neutral probability measures, and the risk premia that link the two.

We implement our estimation and testing procedure on a long time-series of S&P 500 options prices and find strong evidence of self-excitation in jumps. Furthermore, we show that the model is more suitable for fitting 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 also can it 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 for the model parameters.

The model specification with self-exciting jumps, which we focus on, constitutes an alternative parametric approach to explaining the observed differences between the time-series behavior of asset return distributions implied from option prices and the time-series behavior
of asset return distributions derived from the actual price of the underlying asset, through the introduction of a jump risk component. Existing studies\(^2\), 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 at most linked to other state variables (e.g. to the stochastic volatility process). By contrast, in our approach the jump process is linked to its own history - an inherent feature of Hawkes point processes, leading to a realistic and parsimonious approach.

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 sample 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 implied not from a randomly picked pair of options at each time point, but in a uniform way throughout the whole length of the data sample. The parameter estimates and the filtered latent states obtained support the hypothesis of a stochastic jump intensity which would allow for jumps in the S&P 500 index returns to cluster, hence exhibiting a pattern consistent with that of a self-exciting jump process. After a first jump occurs, according to our estimates, the risk premium investors demand to bear the risks of future jumps occurring spikes up and afterwards, if no subsequent jumps occur, it only slowly decreases, the average half-life of an asset price jump shock to the intensity process being approximately two weeks. 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, suggesting that using a well-specified jump process when modeling asset returns is important, particularly during periods of turmoil.

The remainder of the paper is structured as follows: Section 2 describes the properties of the Hawkes self-exciting jump process we employ to model clustered jumps in asset prices, specifies the asset return dynamics and the candidate pricing kernel, and finally provides the option pricing model by deriving in closed-form the conditional characteristic function for our

\(^2\)Among the papers dedicated to this topic are e.g., Bates (2000), Pan (2002), Eraker (2004), Bollerslev and Todorov (2011).
model. Section 3 describes our parametric estimation procedure, proves its consistency, and presents our Monte Carlo results. Section 4 contains our empirical analysis and analyzes the model performance compared to alternative option pricing models. Section 5 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.

2 The Model

2.1 Self-Exciting Jump Processes

In order to model the dynamics of the jump component, we use a Hawkes point process. The Hawkes process,\(^3\) denoted by the pair \((N, \lambda)\), where \(N\) is a counting process with intensity driven by \(\lambda\), is a Markov process. The intensity describes the \(\mathcal{F}_t\)-conditional mean jump rate per unit of time, i.e.:

\[
\begin{align*}
\mathbb{P}[N_{t+\Delta} - N_t = 0|\mathcal{F}_t] &= 1 - \lambda_t \Delta + o(\Delta); \\
\mathbb{P}[N_{t+\Delta} - N_t = 1|\mathcal{F}_t] &= \lambda_t \Delta + o(\Delta); \\
\mathbb{P}[N_{t+\Delta} - N_t > 1|\mathcal{F}_t] &= o(\Delta).
\end{align*}
\] (2.1)

The special feature of the Hawkes process is its self-exciting property, which determines an interdependence between the processes \(N\) and \(\lambda\). This feature is best depicted by the stochastic integral equation which describes the dynamics of the intensity process:

\[
\lambda_t = \bar{\lambda} + \int_{-\infty}^t \delta e^{-\kappa \lambda (t-s)} dN_s. \tag{2.2}
\]

The intensity of \(N\) is determined by \(\lambda\) and, in turn, \(\lambda\) also changes in response to an increase in the counting process \(N\). Jumps feedback into the intensity process which governs the likelihood of future jumps. If we interpret the model in the context of asset price jumps, \(\bar{\lambda} > 0\) represents the intensity until the first (ever) asset price jump occurs, whereas \(\delta e^{-\kappa \lambda (t-s)}\), with \(\kappa > \delta > 0\) measures the (decaying) impact of an asset price jump on the intensity process.

\(^3\)Throughout the paper we refer to probability spaces of the type \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})\) where \(\mathcal{F}_t\) is a right-continuous, complete information filtration satisfying all the usual conditions, see e.g. Protter (2005).
The restriction \( \kappa \lambda > \delta \geq 0 \) ensures that the intensity process is positive and (conditionally) stationary. The impact of the jump decays over time at an exponential rate with speed \( \kappa \lambda \). The compensated process \( N_t - \int_{-\infty}^{t} \lambda_s ds \) is a local martingale. The intensity process (2.2) has the equivalent formulation:

\[
d\lambda_t = \kappa \lambda (\overline{\lambda} - \lambda_t) \, dt + \delta dN_t, \tag{2.3}
\]

This intensity model prescribes a parsimonious way to incorporate the jump history in the time \( t \)-intensity variable \( \lambda_t \). The Hawkes self-exciting point process falls outside the class of Lévy processes. Sample paths from a simulated Hawkes process illustrate typical dynamics of the pair \((N, \lambda)\) in Appendix A.

### 2.2 The Model under \( \mathbb{P} \)

The model describes univariate equity return dynamics and is an adaptation of the classic Bates (2000) model. On a suitably defined filtered probability space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})\), assume that, under the historical measure \( \mathbb{P} \), the log-forward price process \( y_t = \log(F_t) \), the stochastic volatility \( v_t \) and the stochastic intensity \( \lambda_t \) of the point process with counter \( N_t \) have the following dynamics at any given \( t > 0 \):

\[
dy_t = \left( (\eta - \frac{1}{2})v_t + (\mu - \mu^*)\lambda_t \right) dt + \sqrt{v_t}dW^{(1)}_t, \tag{2.4}
\]

\[
dv_t = \kappa_v (v - v_t) dt + \sigma_v \sqrt{v_t} \left( \rho dW^{(1)}_t + \sqrt{1 - \rho^2} dW^{(2)}_t \right), \tag{2.5}
\]

\[
d\lambda_t = \kappa \lambda (\overline{\lambda} - \lambda_t) dt + \delta dN_t. \tag{2.6}
\]

This specification uses the familiar Heston (1993) model for stochastic volatility and a (compound) Hawkes self-exciting point process to model the jumps in the asset return process. \((W^{(1)},W^{(2)})\) are \( \mathcal{F}_t \)-adapted Brownian motions under the physical probability measure \( \mathbb{P} \).

By modeling the log-forward rate instead of log-returns we avoid having to deal explicitly with the modeling of dividend yields.

This model captures two salient features of equity return dynamics: stochastic volatility and price jumps. Stochastic volatility is modeled through the process \( v_t \) defined in (2.5), which

---

\(^4F_t\) denotes the forward price process for a given fixed maturity of a standard forward contract traded on an index or stock.
is also known as a square-root or CIR process from Cox et al. (1985). This volatility process is mean reverting towards a long run mean $\bar{v}$ with a mean reversion rate $\kappa_v$. 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 $\rho$, capturing the leverage effect coined by Black (1976).

Some studies (e.g. Eraker et al. (2003), Eraker (2004)) justify the inclusion of a jump component in the volatility process, to better accommodate volatility dynamics, possibly synchronizing the jumps in volatility with the ones in the stock 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 variations in the model is in part due to the need to better fit volatility dynamics over time. We do not consider any of these variations in the stochastic volatility equation firstly because the square root variance process together with the self-exciting jump process in returns are flexible enough to account for the dynamics of the return distribution over time and, secondly, out of tractability concerns for the estimation approach.

Pertaining to the jump component in our model, $dJ_t^P$ denotes the compound Hawkes jump increment, i.e. $dJ_t^P = Z_t^P dN_t$. It is the result of the product between a serially independent normal random variable $Z_t^P$ governing the jump size and the jump counter instantaneous increment $dN_t$. Conditional on a jump arriving, the forward price jump is $F_t = F_{t-} \exp(Z_t^P)$, where $Z_t^P$ is normally distributed with mean $\mu_j^P$ and standard deviation $\sigma_j$. This implies that the mean relative jump size of the forward price under the physical measure $\mathbb{P}$ is $\mu = \mathbb{E} \left[ \exp(Z_t^P) - 1 \right] = \exp(\mu_j^P + \frac{1}{2}\sigma_j^2) - 1$. The jump times are determined by the counting process $N_t$ with intensity $\lambda_t$, the pair forming the Hawkes process detailed in the previous section. Using a log-normal distribution to model the distribution of asset return jump sizes is a standard choice in the literature concerned with the modeling of asset price jumps, dating back to the seminal paper by Merton (1976).
2.3 Candidate Pricing Kernel and Market Price of Risks

Given the model (2.4)-(2.6), the market setting is rendered incomplete with respect to the underlying asset, the finite number of option contracts available and the money market account, so any pricing kernel in this set-up would not be unique. We restrict our attention to a candidate kernel which keeps the dynamics of the stochastic volatility and the intensity process under the pricing measure \( Q \) within the affine model class.\(^5\) Similar to Pan (2002), Eraker (2004) and Broadie et al. (2007), consider a state price density, \( \xi_t \), of the following form:

\[
\xi_t = \exp \left( -\int_0^t r_s ds \right) \xi_t^D \xi_t^J
\]

\[
\xi_t^D = \exp \left[ -\left( \int_0^t \Gamma_1(s) dW_{1,s}^P + \int_0^t \Gamma_2(s) dW_{2,s}^P \right) \right. \left. - \frac{1}{2} \left( \int_0^t \Gamma_1^2(s) ds + \int_0^t \Gamma_2^2(s) ds \right) \right]
\]

\[
\xi_t^D = \exp \left( \sum_{s^* \leq t} Z_{i,s}^* \right).
\]

In \( \xi_t^D \), the part of the state price density governing the market price of risk for the two Brownian motions, the processes \( \Gamma_1 \) and \( \Gamma_2 \) are defined as:

\[
\Gamma_1(s) = \eta \sqrt{v_s}, \quad \Gamma_2(s) = -\frac{\rho}{\sqrt{1 - \rho^2}} \left( \eta + \frac{\eta}{\sigma_v} \right) \sqrt{v_s}.
\]

The market price of diffusive risks (Brownian shocks) in the log-forward process is pinned down by the \( \eta \) parameter. The risk premium is akin to the risk-return trade-off in the CAPM framework, the larger \( \eta \) is, the higher the premium for diffusive risk is. This risk premium is time-varying, as the volatility process, \( v_s \), changes over time according to its dynamics, the risk premium per unit of time, i.e. \( \eta v_t \), also changes over time. The second risk parameter parameter, \( \eta_v \), determines the market price of diffusive risk in the volatility process, known as volatility risk premium. There is no clear empirical evidence concerning the statistical significance, the magnitude or the sign of the diffusive volatility risk premium parameter \( \eta_v \) stemming from empirical studies on index returns using models with a stochastic volatility state.\(^6\) We therefore opt to constrain the \( \eta_v \) parameter to zero and do not account for a volatility risk premium in the volatility process, i.e., we constrain the volatility process dynamics to be the same under

\(^5\) A definition of the affine jump diffusion model class is given in Duffie et al. (2000).

\(^6\) See e.g. Broadie et al. (2007) for a discussion of the empirical evidence on the size and sign of the volatility risk premia.
both the physical and the pricing measure.

The price density jumps, denoted by \( Z^*_s \), are assumed to be i.i.d. normally distributed with mean \( \mu_j^* \) and variance \( \sigma_j^* \). The stochastic behavior of the state price density jumps governed by their mean and variance together with any correlations they might have with other stochastic variables in the model determine the jump risk timing and jump risk size premiums. We opt to impose that the intensity of jumps to be the same under \( \mathbb{P} \) and under \( \mathbb{Q} \), same as in e.g. Pan (2002), Eraker (2004). This means that state price density jumps at the same time with the underlying asset, i.e., \( \xi^J_t \) takes effect when the underlying jumps as well, the \( s^*_i \) time points denote the times at which the forward price experiences a jump. The \( Z^*_s \) are assumed independent from all other stochastic variables 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, and only correlated ”contemporaneously” with the jump size variable \( Z_s \) in the log-forward price. This is done by imposing the restriction \( \mu_j^* + \frac{1}{2}(\sigma_j^*)^2 = 0 \). The latter restriction keeps the timing of the jumps (hence the intensity process dynamics) the same under both measures, i.e., it restricts the jump risk premiums to be jump-size related and not jump-timing related. We further assume, as is done in Pan (2002), that only the mean of the jump size distribution changes between the risk neutral specification and the physical probability specification of the model.

The jump risk premium originates from the difference between the means of the jump sizes in the log-forward returns, i.e. \( (\mu_j^* \neq \mu_j^Q) \). As a consequence, the mean relative jump size under \( \mathbb{P} \), i.e. \( \mu \) also differs from the mean relative jump size under \( \mathbb{Q} \), i.e. \( \mu^* \). We restrict the jump size variance \( \sigma_j^2 \) to be the same under both measures. In principle, the jump size variance could be different under the two measures, i.e. \( \sigma_j^P \neq \sigma_j^Q \), but this type of premium would be hard to identify in an empirical application. All these choices related to the jump part of the state price density, \( \xi^J_t \) are justified to keep model estimation tractable under both probability measures, albeit the case for a richer specification could be made, this would could only come at the expense of weaker parameter identification. In part the weak identification problem stems from the classical peso problem: as jumps are rare events, pinning down estimates for their jump size distribution is challenging.

Given this set-up of the jump part \( \xi^J_t \) in the state price density, the expected excess return (risk premium) which investors demand in exchange for bearing asset price jump risks

10
is \((\mu - \mu^*)\lambda_t\) per unit of time. This premium is time varying as the latent intensity process \(\lambda_t\) changes over time. Finally, note that if we define the density process to be \(\xi_t \exp(\int_0^t r_s ds)\), we can also define an equivalent martingale measure \(Q\), under which \(W_t^{(i),Q} = W_t^{(i),P} + \int_0^t \Gamma_{i,s} ds\) for \(i = 1, 2\).

2.4 The Model under \(Q\)

Under the equivalent martingale measure associated with the chosen candidate pricing kernel by \(Q\), the log-forward return process has the following dynamics under \(Q\):

\[
\begin{align*}
dy_t &= \left(-\frac{1}{2}\right) v_t dt + \sqrt{v_t} dW_t^{(1),Q} + dJ_t^Q - \mu^* \lambda_t dt \\
\lambda_t &= \kappa_\lambda (\lambda_t - \lambda_t) dt + \delta dN_t.
\end{align*}
\]

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

2.5 Option Pricing

In both its specification under \(P\) and under \(Q\), the model falls into the affine jump diffusion class of models as defined in Duffie et al. (2000). For all affine models, option pricing is efficiently done using Fourier inversion techniques given that the conditional characteristic function of the state vector of stochastic states can be derived. So, in our set-up, we derive the characteristic function of \((y_t, v_t, \lambda_t)^T\) - the state vector, and invert it to find the price of the derivative contract at \(t = 0\) assuming that the initial state vector \((y_0, v_0, \lambda_0)^T\) and the model parameters are known. The computation of the characteristic function can be carried out in closed form up to the solution of a system of ordinary differential equations. Its derivation in the following subsection

11
is based on the results obtained in the general setting of affine jump diffusions developed in Duffie et al. (2000).

2.6 Affine Set-up and Characteristic Function Based Pricing

The system of stochastic differential equations (under $\mathbb{Q}$) which determines the state vector of our model, i.e. $X_t \equiv (y_t, v_t, \lambda_t)^T$ can be written in the following matrix form:

$$
\begin{align*}
\frac{d}{dt} & \begin{pmatrix} y_t \\ v_t \\ \lambda_t \end{pmatrix} = \\
& \begin{pmatrix} -\frac{y_t}{2} - \mu^* \lambda_t \\ \kappa_v (v - v_t) \\ \kappa_\lambda (\bar{\lambda} - \lambda_t) \end{pmatrix} dt + \\
& \begin{pmatrix} \sqrt{v_t} & 0 & 0 \\ \sigma_v \rho \sqrt{\lambda_t} & \sigma_v \sqrt{1 - \rho^2} \sqrt{\lambda_t} & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
& \begin{pmatrix} dW_t^{(1), \mathbb{Q}} \\ dW_t^{(2), \mathbb{Q}} \\ 0 \end{pmatrix} + \\
& \begin{pmatrix} 0 \\ \delta \end{pmatrix} dN_t.
\end{align*}
$$

(2.14)

In the notation of Duffie et al. (2000), this system fits into the class of affine jump diffusions as $\mu(\cdot)$ and $\sigma(\cdot)\sigma(\cdot)^T$ are affine functions of their argument $X_t$, and can be written as:

$$
\frac{dX_t}{dt} = \mu(X_t) dt + \sigma(X_t) dW_t^{3 \times 1} + Z_t^{3 \times 1} dN_t.
$$

(2.15)

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

$$
\phi(u, X_t, T; \theta) = E^\theta [e^{iu \cdot X_t} | \mathcal{F}_t].
$$

(2.16)

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

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

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

\[
\begin{align*}
\dot{\beta}_1 &= 0 \\
\dot{\beta}_2 &= -\frac{\beta_1}{2}(1 - \beta_1) - \kappa_v \beta_2 + \sigma_v \rho \beta_1 \beta_2 + \frac{1}{2} \sigma_v^2 \beta_2^2 \\
\dot{\beta}_3 &= -(e^{\mu_3^Q + \frac{\sigma_3^2}{2}} - 1)\beta_1 - \kappa_\lambda \beta_3 + e^{\beta_1 \mu_3^Q + \frac{\sigma_3^2}{2} + \delta_3} \beta_3 - 1 \\
\dot{\alpha} &= \kappa_v \delta \beta_2 + \kappa_\lambda \bar{\lambda} \beta_3,
\end{align*}
\]
subject to the initial conditions $\beta(T - 0) = (u_1, u_2, u_3)^T$ and $\alpha(T - 0) = 0$.

The proof follows from the application of Proposition 1 in Appendix B of Duffie et al. (2000). This system of ordinary differential equations can be solved using accessible numerical methods (e.g. Runge-Kutta), however a full analytical solution\(^7\) is not possible due to the non-linearities in the differential equation involving $\dot{\beta}_3$.

Having determined how to compute the conditional characteristic function for the log-forward price\(^8\) we use the Fast Fourier Transform approach of Carr and Madan (1999) to price European options. Assuming we denote a set of option contract characteristics, i.e. several pairs of maturities and strike prices, by $C_k = \{T, K\} \in \mathbb{R}^2_+$ which we collect in a matrix $C = (C_1, C_2, \ldots, C_k, \ldots)^T$ and the parameter set by $\theta = (\mu_j^Q, \mu_j^P, \sigma, \eta, \kappa_v, \lambda, \sigma_v, \rho, \kappa_\lambda, \lambda, \delta)$, we define the option pricing function which determines (a vector of) prices for European call\(^9\) options at time $t$:

$$p_t = P(X_t, \theta, C). \quad (2.17)$$

3 Model Estimation

In this section we outline our estimation strategy for the model parameter set $\theta$. The estimation uses a long panel of option prices. Two of the three variables in the state vector are latent, i.e. we only have discrete observations of log-forward returns\(^10\), whereas the volatility and intensity processes are never observable.

One possible approach to estimate the parameters is to minimize the price differences between the market observed call prices and model generated call prices by varying the parameters. This approach, sometimes referred to as ‘calibration’ is often used for option pricing with stochastic volatility models, in applications where the main research objective is to fit a model to a 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 risk premiums or about the longer horizon dynamics of the return distribu-

\(^7\)The first equation is trivially solved by $\beta_1(T - t) = u_1$, while a full analytical solution for the second one which falls into the class of Ricatti differential equations can be found in Duffie et al. (2000).

\(^8\)I.e. by evaluating the system of ODE with the simplified initial condition $\beta(T - 0) = (u_1, 0, 0)$.

\(^9\)Put option prices can be obtained by using the standard put-call parity for European options.

\(^10\)To ensure stationarity, the state vector we use for estimation contains log returns instead of the log price level, as was the set-up for the option pricing exercise.
tion under the physical measure and would not ensure any consistency between the parameters which would be obtained by estimation conducted using time series data of log-returns.

By using the time series of option prices,\textsuperscript{11} we can jointly estimate the $P$ and $Q$ dynamics of the model. We briefly outline the main intuition behind the workings of our estimation method, the full details about each estimation step are presented in the next subsections. If we could observe the full state vector at a set of time points $t_1, t_2, \ldots, t_i, \ldots t_n$, we could use the equations of the model dynamics under $P$ to write down a set of $F_{t_i}$-conditional moment conditions which would also contain and identify all model parameters, including risk premium parameters $\eta$ and $\mu_P$.

We do not observe the full state vector, so, in order to deal with the latent processes $v_t$ and $\lambda_t$ we use market quoted option prices to infer the levels of the latent state processes at the discrete time points at which the option prices are quoted.\textsuperscript{12} This type of estimation approach was used to find model parameter estimates 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 level, i.e. a $\lambda_t = \lambda_1 v_t$, so the state vector only contains the observable log-stock price and the unobservable volatility. Compared to our set-up, the model in Pan (2002) is more parsimonious at the expense of a stronger assumption about the jump intensity process, whose stochastic component is identical to the one generated by the two Brownian components in the volatility process.

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 the two latent factor model set-up such as the one we are in. 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 full parameter vector $\theta$. When implied from option prices at the true parameter vector,

\textsuperscript{11}We use near at-the-money put and call options to determine the price of the forward contract which could serve as an underlying for an option contract written on a forward contract with the same maturity as the option. The time series of log-forward returns is used in lieu of the log-stock returns for practical considerations which are detailed in the next section.

\textsuperscript{12}Alternative iterative estimation procedures like a Kalman filter based approach could be applied in this set-up with latent states, however this would likely result in an efficiency loss by ignoring the specifics of option price non-linearity in the state variables.
say $\theta_0$, the implied-states will be different from when they are implied from a different parameter vector $\theta$. 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 (this only happens for $\theta_0$), we can devise an estimation strategy for the model parameters by ruling out implied state series which are unlikely under our assumed dynamics.

Starting from an initial set of parameters $\theta$, we use the option pricing framework\(^{13}\) to back out $v_{t_i}^\theta$ and $\lambda_{t_i}^\theta$. We evaluate moment conditions derived from the model under $P$ based on the full state vector $(y_{t_i}, v_{t_i}^\theta, \lambda_{t_i}^\theta)^T$ and the choice of $\theta$. We change the choice of $\theta$ and go through the process again until a moment based criterion function is minimized. The parameter set which minimizes the moment based criterion function (for a given sample) is an estimator of the true parameter set $\theta_0$.

### 3.1 Backing out Latent States

Assume that market quoted prices are observed with a regular frequency $\Delta$ at equally spaced\(^{14}\) time points $\{t_1, t_2, \ldots, t_i, \ldots, t_N\}$ for the log-forward price $y_{t_i}$ and for European-style derivative contracts $\{p_{t_i,k}\}, k = 1, \ldots, m$ with $m \geq 2$ ($k$ indexes all strike-price ratio and maturity combinations for option contracts which have a price recorded 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 $X_{t_i} = (y_{t_i}, v_{t_i}, \lambda_{t_i})^T$ and the implied state vector by $X_{t_i}^\theta = (y_{t_i}, v_{t_i}^\theta, \lambda_{t_i}^\theta)^T$, where the $\theta$ superscript 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 $\Theta$ being a compact parameter space. Recall that $P(X_{t_i}, \theta, C)$ is the option pricing function defined previously in equation (2.17). We define the inverse mapping: $f(p_{t_i}, \theta, C) : \mathbb{R}^m_+ \times \Theta \times \mathbb{R}^{2 \times m} \to \mathbb{R} \times \mathbb{R}^2_+$, which recovers the latent states from option prices, through the following equation:

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

(3.1)

The above equation states that the mappings are defined such that when the latent states

\(^{13}\)We use the superscript $\theta$ to highlight the explicit dependence of the implied latent state variables on the parameter set $\theta$.

\(^{14}\)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.
are recovered from a set of option prices and then plugged back in the option pricing function the result is the set of option prices which we started with, i.e. the option pricing function as a mapping of the state variable is invertible. Part of the estimation procedure, the latent states are repeatedly implied at parameter values different from the true parameter set until we arrive at the true parameter set. We formalize the link between the true state vector $X_t$ and the state vector $X^\theta_t$ which is obtained at the different $\theta$ parameter set by using the the state implying function $f(\cdot)$:

$$X^\theta_t = f(P(X_t, \theta_0, C), \theta, C).$$

(3.2)

By assuming the inversion is well defined, we have that $X^\theta_0 = f(P(X_t, \theta_0, C), \theta_0, C) = X_t$, or in other words the true levels of the latent states can be implied when using the true parameter set $\theta_0$. In simulation exercises we conducted, the levels of simulated states were always backed out with a high degree of numerical accuracy from option price panels generated based on the respective simulated state series paths. The inversion was well behaved at all parameter set values we tried, 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.

Each sample period there are more 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 latent volatility state needed for the estimation. Pastorello et al. (2003) proposes introducing a pricing error with a stationary 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. (2012) impose weak in-fill asymptotic assumptions on the option pricing error distribution. In the same spirit, we make a set of assumptions about the pricing errors in our state-implying context.

Given that the Black Scholes implied volatility function is a monotonic transformation of dollar option prices, it makes no difference whether the price series $p_t$ is expressed in dollar amounts or implied volatility points. We use the latter as input in the state-backing out function. In our context it 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 moneyness.

---

15 This assumption is implied by Assumption 2 in Appendix B.
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 the prices and prevents the over-weighting of in-the-money options.\(^{16}\)

To obtain the standardized Black-Scholes implied volatilities at all the combinations of option contract maturities \(T_l\) and moneyness levels \(K_m/F\) we use a Gaussian kernel (Cont et al., 2002):

\[
\text{BSvol}_t(T_l, K_m/F) = \frac{\sum_{k=1}^{n} \text{BSvol}_t(T_k, K_k/F) u(T_l - T_k, K_m/F - K_k/F)}{\sum_{q=1}^{n} u(T_l - T_k, K_m/F - K_k/F)}. \tag{3.3}
\]

Here, \(k\) denotes the index of all the options available in a given sample day \(t_i\), while \(l\) and \(m\) index the maturities and moneyness levels. The smoothing function is:

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

In our empirical application, the bandwidth parameters \(h_1\) and \(h_2\) were chosen through trial and error 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.

If option prices were observed without error, then, for each period, two option prices would be enough to back out the level of the two latent state variables, the stochastic volatility and the self-exciting jump intensity. We make a series of assumptions about the pricing errors, detailed in Appendix B. Under these assumptions, by using in-fill asymptotic arguments the pricing errors would vanish after the kernel smoothing procedure for the implied volatility surface is applied. Because in practice we always smooth the implied volatility surface using a finite number of option prices, 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

\(^{16}\)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 backing-out routine.
At time $t_i$, we minimize the (sum of squared) differences between the Black-Scholes implied volatilities of the standardized option prices from the kernel smoothing and the Black-Scholes implied volatilities from the option prices obtained using the model pricing approach at the assumed parameter values:

$$X^\theta_{t_i} = \min_{V_i, \lambda_i} \sum_k \left( \text{BSvol}(p_{t_i,k}) - \text{BSvol} \left( P(X^\theta_{t_i}, \theta, C) \right) \right)^2. \quad (3.5)$$

In our empirical application, for each sample day we considered three maturities $T_l \in [0.1, 0.5, 1]$ (expressed in years\footnote{We follow the convention that 1 year is made up of 252 trading days.}) and a set of relative moneyness levels $K_m / F \in [0.8, 0.9, 1, 1.1, 1.2]$. The downside of using more option contracts is the increase in computational cost. 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 a different parameter set until the criterion function is minimized. The choice of criterion function is detailed in the next subsection.

### 3.2 Estimation Procedure

Given that the conditional characteristic function of the state vector can be computed, the likelihood function of our model could be recovered through successive applications of multidimensional Fourier integration. However, approximating the likelihood function in this way is computationally expensive and impossible to implement in a practical application. As a feasible alternative, we develop an approach based on GMM. Optimal moment conditions are usually inferred from the first order conditions of a model’s likelihood function, 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 first order conditions of the likelihood function. We follow this route and use $\phi^\mathbb{P}(\cdot; \theta)$, the conditional characteristic function of the state vector under the physical probability measure, to construct moment conditions. The model set-up under the two probability measures allows for the estimation of the model parameters which characterize the $\mathbb{P}$ dynamics together with the parameters which characterize the $\mathbb{Q}$ dynamics, as the model under $\mathbb{P}$ also contains all the
parameters in the model under $\mathcal{Q}$. We derive the conditional characteristic function under $\mathbb{P}$ in the same way we derived the conditional characteristic function under the risk neutral measure in the previous proposition and use it to build moment conditions.

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

$$h(r, s, X_{t_i}, X_{t_{i+1}}, \theta) \equiv m(r, X_{t_i}) \left( e^{sX_{t_{i+1}}} - \mathbb{P}(s, X_{t_i}, \Delta; \theta) \right).$$

(3.6)

The real vector $s$ denotes the argument at which the conditional characteristic function is evaluated, while $m(r, X_{t_i})$ denotes an arbitrary instrument with a real argument vector $r$. Following from the definition (2.16), the set of moment conditions (3.6) is a martingale difference series. Two choices have to be made when using the set of moment conditions based on the conditional characteristic function: the choice of argument $u$ and the choice of instruments $m(r, X_{t_i})$. Singleton (2001) shows that using a finite set of vectors $\{s_1, s_2, \ldots\}$ such that each of the components are equally spaced leads to a consistent estimator, even if such a grid contains only few distinct values for each component. In practice, although efficiency gains are obtained when the grid is refined, the inversion of the sample moment condition covariance is numerically difficult as it becomes singular in practical implementations. For the instrument set, Singleton (2001) proposes to compute $m(r, X_{t_i})$ as an optimal instrument from the Hansen (1985) framework. Carrasco et al. (2007) investigate the efficiency of this type of estimator using a continuum of $s$ values, and show that further instrumenting the continuum of moment conditions with a continuum of instruments of the type $m(r, X_{t_i}) = e^{rX_{t_i}}$ leads to considerable efficiency gains in the estimation procedure.$^{18}$

The efficient estimation procedure using a continuum of moment conditions developed in Carrasco and Florens (2002) and Carrasco et al. (2007) is mostly dedicated to estimation contexts in which the full state vector is observed. As in our set-up and in many other asset pricing contexts, latent state variables are usually contained in the state vector to allow for the modeling of richer dynamics for asset return distributions. A solution in a wide range of

---

$^{18}$The intuition behind this functional form for the instrument is that, when evaluated over a continuum of $r$ values, it spans the partial derivative of the log transition density of $X_t$ w.r.t. the parameter set $\theta$, which is the optimal, yet infeasible, instrument in this set-up, see Singleton (2001) and Carrasco et al. (2007).
estimation problems with latent state variables, discussed in Carrasco et al. (2007), is to replace these latent state variables in the state vector with simulated paths in the estimation routine. There are many asset pricing instances in which information about the latent states can be extracted from market prices of contingent claims linked to the distribution of the state vector. We explore this alternative route and 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 is observable. Introducing the notation \( \tau \equiv (r, s) \), the time \( t_i \) sample moment condition in our implied state set-up is:

\[
h_{t_i} (\tau, \theta) \equiv h(\tau, X_{t_i}^\theta, X_{t_i+1}^\theta, \theta) = e^{irX_{t_i}^\theta} \left( e^{isX_{t_i+1}^\theta} - \phi(s, X_{t_i}^\theta, \Delta; \theta) \right).
\]

(3.7)

Denote the sample mean of the moment condition evaluated at \( \tau \) by:

\[
\hat{h}_n (\tau; \theta) = \frac{1}{n} \sum_{i=1}^{n} h(\tau, X_{t_i}^\theta, X_{t_i+1}^\theta; \theta).
\]

(3.8)

The estimator is based on the convergence of the moment conditions sample mean evaluated to \( \mathbb{E}^\theta (h_t(\tau; \theta_0)) = 0 \) for any \( \tau \). Following Carrasco et al. (2007), we employ a continuum of moment conditions indexed in the \( \tau \)-argument by introducing the Hilbert space equipped with the following inner product, for any two complex-valued functions \( f, g \):

\[
\langle f, g \rangle = \int_{\mathbb{R}^{\dim \tau}} f(\tau) \overline{g(\tau)} \pi(\tau) d\tau,
\]

(3.9)

where \( \overline{g(\tau)} \) denotes the complex conjugate of \( g(\tau) \) and \( \pi(\tau) \) denotes 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 distribution function. The Gaussian density is a computationally convenient choice because of the accessible quadrature methods for multidimensional integration which are available when implementing the estimator in practice.

\[19\] A formal definition of the Hilbert space is given in Assumption 3, in the appendix section.
Using a continuum of values for $\tau$, one can obtain a first step consistent GMM estimator by solving:

$$\hat{\theta}_{1st}^n = \arg\min_{\theta} \| \hat{h}_n(\tau; \theta) \| = \arg\min_{\theta} \int \hat{h}_n(\tau; \theta)\overline{h}_n(\tau; \theta)\pi(\tau)d\tau. \quad (3.10)$$

We find a first step estimator for our data sample by successively backing out the latent states from option prices and computing the integral in (3.10) until the latter is minimized.

**Proposition 2:** Under Assumptions 1-4 (detailed in Appendix C1) the estimator $\hat{\theta}_{1st}^n$ converges to $\theta_0$ in probability as $n \to \infty$.

The assumptions presented in the appendix ensure that the conditions required for the proof of consistency are also met in our extended set-up in which two of the states are backed out from option prices. We discuss the validity of the assumptions mentioned in Proposition 2 for the model with self-exciting jumps which we consider in our application, see Appendix C3. In the spirit of Carrasco et al. (2007), we also compute a second step GMM estimator which uses the first step estimator as an input to calculate $K$, the covariance operator applied to our moment conditions:

$$K f(\tau_1) = \int k(\tau_1, \tau_2)f(\tau_2)\pi(\tau_2)d\tau_2, \quad (3.11)$$

$$\text{with} \quad k(\tau_1, \tau_2) = E^{\theta_0} \left( h_n(\tau_1; \theta_0)\overline{h}_n(\tau_2; \theta_0) \right). \quad (3.12)$$

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, adapted to the context in which a continuum of moment conditions is used. The second step estimator then becomes:

$$\hat{\theta}_{2nd}^n = \arg\min_{\theta} \| (K_n^{\alpha_n})^{-1/2}\hat{h}_n(\tau; \theta) \|. \quad (3.13)$$

The superscript $\alpha_n$ denotes the regularization of the inverse of the covariance operator $K$, which is needed to ensure a stable solution for the invertibility problem.\(^{20}\)

**Proposition 3:** Under Assumptions 1-7 (detailed in Appendix C4) the estimator $\hat{\theta}_{2nd}^n$ converges to $\theta_0$ in probability as $n \to \infty$.

\(^{20}\)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. (2007).
3.3 Practical Considerations and Monte Carlo Evidence

Because of the additional latent-state implying step procedure which is based on the numerical option pricing routine, analytic results about the efficiency of the estimator cannot be derived. This immediately raises two follow-up issues. First, a theoretical justification for the second step estimator, such as proving an efficiency gain for the estimator, cannot be made. We 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 should lead to an efficiency gain, albeit the covariance operator we consider here ignores the latent-state backing out. Second, we need to find an approximation for parameter standard errors. To this avail, we resort to obtaining an approximation of the standard errors by inverting the numerically approximated Hessian of the second step criterion function.

To assess the sample performance of the proposed estimation approach we conduct a Monte Carlo exercise. We simulate state vector series from the SVHJ model 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.

We simulated state vector levels by discretizing their continuous-time dynamics. Each simulated sample contained 9 option prices (maturities 0.1, 0.5 and 1 year and moneyness ratios of 95%, 100% and 105%) for each of 500 weekly observations. Therefore, in the $T$ dimension, the simulated option panel contained 500 observations, while the cross-section dimension was constant for each $t$ and set to 9 option prices. We ran the estimation for 100 simulated option panels, thus obtaining 100 sets of parameter estimates. This number of samples is small compared to the typical number of replications used in Monte Carlo exercises, but, given the high computational costs, it was a feasible implementation which gives some intuition about the small sample behavior of the estimator. The table below presents the fixed (true) values for the parameter set and the mean and Monte Carlo sample standard deviation as well as some quantiles for the estimator distribution.

---

21Euler discretization, see e.g, Lord et al. (2010) for details.
Table 1: Monte Carlo Simulations

<table>
<thead>
<tr>
<th></th>
<th>$\mu_j^P$</th>
<th>$\mu_j^Q$</th>
<th>$\sigma_j$</th>
<th>$\eta$</th>
<th>$\kappa_v$</th>
<th>$\tau$</th>
<th>$\sigma_v$</th>
<th>$\rho$</th>
<th>$\kappa_\lambda$</th>
<th>$\lambda$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>-5%</td>
<td>-14%</td>
<td>6%</td>
<td>2.40</td>
<td>4.80</td>
<td>0.01</td>
<td>0.22</td>
<td>-0.60</td>
<td>18.00</td>
<td>0.30</td>
<td>16.5</td>
</tr>
<tr>
<td>MC Mean</td>
<td>-7.51%</td>
<td>-14.13%</td>
<td>5.31%</td>
<td>2.87</td>
<td>5.31</td>
<td>0.01</td>
<td>0.23</td>
<td>-0.61</td>
<td>22.51</td>
<td>0.33</td>
<td>16.26</td>
</tr>
<tr>
<td>MC S.D.</td>
<td>3.03%</td>
<td>1.57%</td>
<td>1.96%</td>
<td>0.83</td>
<td>0.92</td>
<td>0.001</td>
<td>0.03</td>
<td>0.08</td>
<td>4.46</td>
<td>0.03</td>
<td>2.65</td>
</tr>
<tr>
<td>10%</td>
<td>-11.16%</td>
<td>-16.33%</td>
<td>3.37%</td>
<td>1.86</td>
<td>4.30</td>
<td>0.009</td>
<td>0.19</td>
<td>-0.72</td>
<td>16.52</td>
<td>0.28</td>
<td>13.43</td>
</tr>
<tr>
<td>25%</td>
<td>-9.48%</td>
<td>-14.91%</td>
<td>4.88%</td>
<td>2.25</td>
<td>4.71</td>
<td>0.009</td>
<td>0.21</td>
<td>-0.68</td>
<td>18.87</td>
<td>0.31</td>
<td>14.64</td>
</tr>
<tr>
<td>50%</td>
<td>-7.87%</td>
<td>-14.12%</td>
<td>5.90%</td>
<td>2.88</td>
<td>5.18</td>
<td>0.01</td>
<td>0.23</td>
<td>-0.62</td>
<td>22.84</td>
<td>0.33</td>
<td>16.17</td>
</tr>
<tr>
<td>75%</td>
<td>-5.34%</td>
<td>-13.16%</td>
<td>7.34%</td>
<td>3.46</td>
<td>6.16</td>
<td>0.012</td>
<td>0.24</td>
<td>-0.54</td>
<td>25.94</td>
<td>0.35</td>
<td>17.88</td>
</tr>
<tr>
<td>90%</td>
<td>-3.73%</td>
<td>-11.94%</td>
<td>8.15%</td>
<td>3.97</td>
<td>6.61</td>
<td>0.013</td>
<td>0.26</td>
<td>-0.51</td>
<td>28.06</td>
<td>0.38</td>
<td>19.54</td>
</tr>
</tbody>
</table>

4 Empirical Application

4.1 Data description

To apply the estimation procedure we used options data on the S&P 500 index from the Option Metrics database. The data-set we used 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 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. We keep an out-of-sample period to benchmark the option-pricing performance of our model. 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 mentioning that the applicability of our model is general so we expect that the model achieves a good fit during non-turbulent periods as well as turbulent ones.

As the data-set contains prices for a large number of option contracts, it would be compu-
tationally over-demanding to back out the implied states and compute model option prices for every sample day. We follow inter alia Pan (2002) and Johannes et al. (2009) and use weekly options data. The day of the week we selected to include in our estimation sample was Wednesday as this is a typical choice in the literature. Using panels of options data to estimate continuous time-stochastic process poses some specific challenges which we detail henceforth. The readers familiar with the issues surrounding options data-sets can skip to the next subsection.

Given that the data set covers a long time period which likely spans more than one business cycle, one must find a way to account for time-variation in interest rates and time-variation in dividend yields. This is an important modeling step as the interest rate and dividend yield levels are parameters which appear in the drift specification of the log-stock 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-stock 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, therefore also circumventing this latter problem. Other approaches to account for these options-specific issues have been taken in the literature. Pan(2002) models the dividend yield and the interest rate separately, each as an affine diffusion and then uses them together with the log-stock returns in the option pricing routine. As we do not have access to reliably synchronized stock and option price data we did not follow this route.

Our estimation 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 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. We use the midpoint between the bid and the ask as an approximation for each option’s true price. We then discard some of the
(outlying) 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 \( T \in [20 \text{ days}, 300 \text{ days}] \), the mid-price > 0.1 (10 USD cents), as 0.1 is the minimum tick on the CBOE and the strike/price ratio \( \in [50\%, 150\%] \). We apply the same criteria for put and call options.

From the selected options we build a standardized data-set. To do that, for each sample day and for each maturity available (e.g. \( \tau \)) in that sample day we imply the forward\(^{22}\) price \( F_{t,\tau} \) for the forward contract with maturity \( \tau \) using the (closest to) at-the-money pair of call & put prices in the put-call parity for options written on a forward contract:

\[
\text{Call}_t + Ke^{-r_t,\tau \tau} = \text{Put}_t + F_{t,\tau}e^{-r_t,\tau \tau}.
\] (4.1)

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. e.g, 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 illiquidity of these contracts) with the recalculated ones which should be a more precise approximation of the true price at which (deep) in-the-money options would trade at. 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 all available put and call contracts in the original data-set.

Finally, in our empirical application based on this adjusted panel of option prices we use the kernel smoothing procedure discussed in the previous section to obtain the standardized sets of Black-Scholes implied volatilities which constitute the data-input for the estimation procedure.

\(^{22}\)When modeling log-forward returns \( y_t \) we assume that it does not matter which contract maturity we base our log-forward return calculation on, as cash-and-carry arbitrage should ensure that all forwards have the same dynamics. This relies on the simplifying assumption that the interest rate term structure remains unchanged until the next sample week. While not realistic, this assumption is reasonable and provides tractability for our estimation procedure as it allows us to forego the complications of explicit interest rate term structure modeling.
4.3 Parameter Estimates & Implied States

The GMM estimates for the model parameters and standard errors in brackets using weekly (Wednesday’s) options data on the S&P 500 index, January 1996 to December 2009, are:

Table 2: Estimated model parameters

<table>
<thead>
<tr>
<th>$\mu_j^p$</th>
<th>$\mu_j^Q$</th>
<th>$\sigma_j$</th>
<th>$\eta$</th>
<th>$\kappa_v$</th>
<th>$\tau$</th>
<th>$\sigma_v$</th>
<th>$\rho$</th>
<th>$\kappa_\lambda$</th>
<th>$\lambda$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.86%</td>
<td>-13.68%</td>
<td>6.63%</td>
<td>2.37</td>
<td>4.76</td>
<td>0.011</td>
<td>0.225</td>
<td>-0.61</td>
<td>18.16</td>
<td>0.326</td>
<td>16.62</td>
</tr>
<tr>
<td>(1.48%)</td>
<td>(1.69%)</td>
<td>(2.94%)</td>
<td>(1.4)</td>
<td>(1.25)</td>
<td>(0.002)</td>
<td>(0.08)</td>
<td>(0.24)</td>
<td>(1.65)</td>
<td>(0.017)</td>
<td>(2.51)</td>
</tr>
</tbody>
</table>

All parameter estimates are statistically significant at conventional confidence levels, with the exception of the parameter governing the volatility risk premium $\eta$. The long run average volatility as deduced from the estimate above is $\sqrt{\nu_t} \approx 10.5\%$. Our model features two sources which generate volatility 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 volatility of the asset $\sqrt{\nu_t}$ (together with the weekly returns for the S&P 500 index) and the stochastic jump intensity (with the level of the S&P 500 index in the background). The jump intensity quickly increased around the past dates of turbulent events such as the 1998 Russian Ruble crisis, the dot-com bubble in the 2000’s and the more recent financial crisis which began in late 2008. The estimates for $\hat{\kappa}_\lambda = 18.16$ and $\hat{\delta} = 16.62$ suggest an intensity process with relatively persistent jumps in its level as a result of asset price jumps occurring. Given the above values for $\kappa_\lambda$ and $\delta$, the half-life a shock to the intensity\textsuperscript{23} generated by an asset price jump is approximately two weeks. The unconditional expected intensity is $\lambda \equiv \kappa_\lambda \bar{\lambda}/(\kappa_\lambda - \delta) \approx 3.8$, meaning that unconditionally one would expect a jump in the S&P 500 index level to occur every quarter.

\textsuperscript{23}Conditional on no other jumps occurring in the meanwhile.
4.4 Risk Premiums

Stemming from the candidate pricing kernel, two time-varying risk premiums are specified in the model: a volatility premium and a jump risk premium. At any point in time, the volatility premium per annum is given by $\eta V_t$. Evaluated at the long run mean volatility level $\bar{\sigma} = 0.011$ this would imply an average long run volatility premium of approximately 2.61% per year. The jump risk premium is determined by the difference between the relative jump sizes of asset jumps under the two equivalent measures, i.e. expressed as $(\mu - \mu^*)\lambda_t$ per unit of time. Figure 2 plots the volatility risk premium, the jump risk premium as well as the total risk premium over time. In the self-exciting jump framework, the jump risk premium increases abruptly in times of turmoil, an empirical finding which is consistent with the impression that investors exhibit increased fear of future jumps after jumps occur and hence they will only be willing to bear this risk if compensated.

Insert Figure 2 here

4.5 Model-Generated Implied Volatility Surfaces

One of the primary applications of continuous time stochastic models for asset returns is option pricing. The self-exciting feature of the jump process together with the mean reversion feature in stochastic volatility should render our model more flexible in accommodating the day-to-day dynamics of the implied volatility surface. The flexibility stems from the two stochastic state variables which vary continuously and determine the shape of the implied volatility surface such that the improved fit is achieved without having to re-estimate the model parameters every period as is the case with many other option pricing models which rely only on a stochastic volatility state variable to generate the dynamics of the volatility surface. To illustrate this, we compare the option pricing performances of several stochastic volatility models with jumps. The models we consider are: stochastic volatility with Poisson jump (labeled SVJ), stochastic volatility with volatility driven jump intensity (labeled SVVJ) and the stochastic volatility with Hawkes self-exciting jumps (labeled SVHJ). All models share the same set-up for the candidate

---

24 The specifications of the SVJ and SVVJ models under the physical measure $P$ and the risk neutral measure $Q$ are presented in Appendix D.
pricing kernel and are estimated using the same procedure, so the focus of the comparison is the impact of the jump component specification on overall option pricing performance.

The SVJ and SVVJ models have only one latent state, the stochastic volatility. By backing out the stochastic volatility levels from quoted derivatives prices we construct a set of state-vector observations which we further use to compute moments based on the conditional characteristic functions of these benchmark models. The data-sample we employed for the estimation of the benchmark models was the same one we used for the SVHJ model, i.e. S&P 500 options covering the time span from January 1996 to December 2009, transformed into standardized sets of implied volatilities as described in the data selection procedure section. The parameter estimates for the SVJ and SVVJ models are presented in Appendix D.

To evaluate option pricing performance we take into consideration several pricing error metrics\footnote{A typical option pricing benchmarking exercise can be found in Bakshi et al. (1997).} and look at the average and root mean squared differences between market observed implied volatilities and model generated implied volatilities during the in-sample period (January 1996 to December 2009) and the out-of-sample period (January 2010 to August 2013). The latent state(s) for each model are backed out\footnote{Using the parameter estimates previously obtained for each of the models via the GMM procedure.} for every sample period from the respective day’s sub-sample of derivatives contracts used for estimation. With the state(s) backed out, the remaining points on the implied volatility surface which are quoted in that day are also evaluated using each of the models. In other words, each of the models is used as an interpolation tool for the implied volatility surface, i.e. assuming some of the surface in that day is revealed, the latent state(s) is (are) backed out in order to price ‘new’ derivatives contracts. These new prices are compared with the actual prices recorded in that sample day. The pricing differences recorded for each of the models are then compared to reveal which model is better at fitting the implied volatility surface.

Figures 3 and 4 show the mean and, respectively, the root mean squared error of the pricing errors expressed in implied volatility percentage points. The model with self-exciting jump intensity leads to the smallest pricing errors, albeit short maturity deep out of the money options are still on average somewhat underpriced by the SVHJ model. The SVVJ model
is second in terms of pricing performance, also underpricing short maturity, deep out of the money options, whilst overpricing the remaining maturity and moneyness 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 together with the stochastic volatility is evident in this set-up and among the two stochastic intensity models benchmarked, the SVHJ results in smaller option pricing errors achieving a better fit for market implied volatility surfaces.

5 Conclusions

For option pricing purposes and the study of risk premiums, we proposed using a fully parametric model for asset returns which is more flexible in accommodating price jump patterns than the classic Poisson jump diffusions or than jump diffusion models in which the jump intensity is completely determined by another stochastic state variable such as the volatility process. After assuming a parametric form for the state price density we derived a closed form approach to option pricing based on this model and devised a way to estimate its parameters using the time series of option prices. After implementing this procedure for the S&P 500 index options we find evidence in favor of modeling the jump process as a (Hawkes) self exciting process.
List of Figures

Figure 1: Time series of option-implied latent states
Figure 2: Risk premiums (% per annum)

Figure 3: Mean Pricing Errors (implied volatility % points)
Figure 4: Root Mean Squared Pricing Errors (implied volatility % points)
Appendix A
Simulated paths from a Hawkes jump process

Sample simulated series with parameters:

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

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

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

B0. Assumptions for the In-Fill Asymptotic Behavior of Option Price Errors

We assume that, as the cross-section dimension of the panel option quotes increases (in-fill asymptotics), the price differences between the true Black-Scholes implied volatilities and the Black-Scholes volatilities implied at the mid-price ‘wash out’ due to the kernel smoothing procedure applied.

As in Andersen et al. (2012), we assume that the implied volatility prices (the superscript in $\text{BSVol}_t^0(T_q,K_q/F)$ is used to denote the true implied volatility observed without error) are observed with an additive error. Then, we assume that the implied volatility pricing errors have a zero mean and that a version of a weak law of large numbers for weighted sums of random variables applies (e.g, see Padgett and Taylor 1978).

Assumption 0:

$\text{BSVol}_t^0(T_q,K_q/F) = \text{BSVol}_t^0(T_q,K_q/F) + \epsilon_t(T_q,K_q/F)$ the observation error is additive.

$E\left[\epsilon_t(T_q,K_q/F)\right] = 0, \forall t$ and $\forall \{T_q,K_q/F\}$, i.e. price errors do not bias prices.

A law of large numbers holds such that:

$$\frac{\sum_{q=1}^{n} \epsilon_t(T_q,K_q/F)k(T_i - T_q, K_j/F - K_q/F)}{\sum_{q=1}^{n} k(T_i - T_q, K_j/F - K_q/F)} \xrightarrow{P} 0.$$ 

B1. Assumptions for Estimator Consistency

Assumption 1:

$X_t$ is a stochastic process, a $p \times 1$ vector of random variables. The distribution of $(X_1,X_2,\ldots,X_T)$ is indexed by a finite dimensional parameter vector $\theta \in \Theta \subset \mathbb{R}^q$ where $\Theta$ is compact.

Assumption 2:

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

$h_t(\tau,\theta)$ is a continuous function of $\theta$.

Assumption 3:

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

$L^2(\pi)$ is the Hilbert space of complex valued functions that are square integrable with respect
to $\pi$:

$$L^2(\pi) = \{g : \mathbb{R}^d \rightarrow \mathbb{C} \text{ where } \int |g(t)|^2 \pi(t) dt < \infty\}.$$  

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

$$\langle f, g \rangle = \int_{\mathbb{R}^d} f(\tau) \overline{g(\tau)} \pi(\tau) d\tau,$$

$$\|f\| = \langle f, f \rangle^{1/2}.$$  

**Assumption 4:**

The equation $E_{\theta_0}[h_t(\tau, \theta)] = 0$, for $\forall \tau \in \mathbb{R}^d$, $\pi$–almost everywhere has a unique solution $\theta_0 \in \Theta$.

**Assumption 5:**

$Q_T \rightarrow Q$, i.e., $\| \hat{h}_n(\tau; \theta) \| \rightarrow \| E_{\theta_0}[h_t(\tau, \theta)] \|$ almost surely uniformly on $\Theta$.

**Assumptions for the Consistency of the 2nd Step Estimator**

**Assumption 6**

Assumptions about $K$, the asymptotic covariance operator of $\sqrt{T}h_T(\theta_0)$:

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

(ii) $E_{\theta_0}[h_t(\cdot, \theta)] \in \mathcal{H}(K)$, $\forall \theta \in \Theta$.

**Assumption 7**

(i) $h_t(\tau, \theta) \forall \tau \in \mathbb{R}^d$ is differentiable w.r.t $\theta$.

(ii) $E_{\theta_0}\left[\sup_{\theta \in \Theta}\|\nabla_{\theta}h_t(\cdot, \theta)\|\right] < \infty$.

(iii) $K$ is a Hilbert-Schmidt operator. In the current set-up for $h_t(\tau, \theta)$, this implies the following conditions on $h$:

- $h_t(\tau, \theta)$ is a bounded function, $\forall \tau \in \mathbb{R}^d$ and $\forall \theta \in \Theta$.

- The $\alpha(\cdot)$-mixing coefficients are summable.

**B2. Proof of Proposition 2**

Given Assumptions 1 to 5, the proof of Proposition 2 follows from Theorem 3.3 in Gallant and White (1988) (henceforth referred to as GW1988).

The correspondence between Assumptions 1 to 5 and the assumptions made in GW1988
is as follows: Assumption 1 contains the GW1988 DGP assumption, while Assumptions 2 and 3 correspond to the GW1988 optimand assumption. A4 corresponds to GW1988 assumption on the identifiable uniqueness (of the criterion minimizing \( \theta \)), while A5 corresponds to GW1988 the uniform convergence assumption. We discuss the validity of Assumptions 1 to 5 for the SVHJ model in Appendix B4.

**B3. Proof of Proposition 3**

The proof of Proposition 3 is similar to that given for the proof of estimator consistency given for the first step estimator, with two extra steps needed to establish asymptotic properties for the covariance operator \( K \). Under the conditions given in Assumptions 1 to 5, Theorem 4.i in in Carrasco and Florens (2000) helps to conclude that:

\[
|K_T - K| \xrightarrow{P} 0.
\]

Using Lemma B.2 in the appendix of Carrasco et al (2007), we can conclude that:

\[
\left| (K_T^{\alpha T})^{-1/2} - (K^\alpha T)^{-1/2} \right| \xrightarrow{P} 0.
\]

Assumption 7 is needed to prove the latter. Then by A6 together with A1 to A5 the conditions for required for Theorem 3.3 in Gallant and White (1988) are met and the weak consistency result for the second step estimator is obtained.

**B4. Assumptions’ validity for the SVHJ model**

This appendix presents some proofs which support the validity of the assumptions needed for the estimator’s asymptotic consistency in the case of the SVHJ model. Not all the assumptions can be supported by proofs as some of the results for the SVHJ model are not available in an analytic closed form. As the conditional characteristic function for the model does not have an analytic solution, some of the derivations which would require the full knowledge of this function can not be shown, but are expected to hold.

**Assumption 1** - \( \{X_t\} \) is an adapted stochastic process with dynamics defined by equations (2.4)-(2.6). As the conditional characteristic function is a function for the model under \( \mathbb{P} \), which
is used in the estimation procedure, is a function of the full parameter set \( \theta \) (containing the risk premium related parameters), the transition density is also a function of the full parameter set.

**Assumption 2** - The properties of the \( h_t(\tau, \cdot) \) function depend on the properties of the state-implying function \( X_t^\theta = f(P(X_t, \theta_0, C), \theta, C) \), as the latter serves as input in the moment condition function. As the pricing function is not available in closed form, the properties of its inverse cannot be established analytically. The inverse function was well behaved in all the simulations and empirical applications. The continuity of the moment condition function \( h_t(\tau, \theta) \) in \( \theta \) hinges on the continuity of the latent-state implying function \( X_t^\theta = f(P(X_t, \theta_0, C), \theta, C) \) in \( \theta \). Under additional mild assumptions which are required for the inverse function theorem to hold, such as the continuity of the partial derivatives of \( f(P(X_t, \theta_0, C), \theta, C) \) w.r.t. to the latent states, the state-implying function is a continuous function of the parameter set \( \theta \in \Theta \).

**Assumption 3** - The following proof shows that the moment condition function is bounded and hence square integrable.

As the conditional characteristic function is bounded, i.e., \( |\phi(\theta, s, X_t^\theta, \Delta; \theta)| \leq 1, \forall s \in \mathbb{R}^d \):

\[
|h_t(\tau, \theta)| = |e^{irX_t^\theta}e^{isX_t^\theta} - \phi(s, X_t^\theta, \Delta; \theta)| \leq |e^{i(rX_t^\theta + sX_t^\theta + 1)}| + |e^{irX_t^\theta}\phi(s, X_t^\theta, \Delta; \theta)| \leq 2
\]

In our applications we set \( \pi(\cdot) \) to be the Gaussian multivariate density, which meets the conditions stated in Assumption 3.

**Assumption 4** - The first implication of the Assumption 4 is true by virtue of the way in which the state-implying is defined, i.e., according to \( X_t^\theta_0 = f(P(X_t, \theta_0, C), \theta_0, C) = X_t \) and therefore the moment condition evaluated at \( \theta_0 \) becomes:

\[
h_t(\tau, \theta_0) = e^{irX_t^\theta_0}(e^{isX_t^\theta_0} - \phi(s, X_t^\theta_0, \Delta; \theta_0)) = e^{irX_t}(e^{isX_t+1} - \phi(s, X_t, \Delta; \theta_0))
\]

Evaluated at the true parameter vector \( \theta_0 \), the conditional characteristic function is then:

\[
\phi(s, X_t, \Delta; \theta_0) = \mathbb{E}^\theta_0\left[e^{isX_t+1}|X_t\right]
\]
The expectation of the moment condition evaluated at the true parameter vector $\theta_0$ is:

$$E^{\theta_0} [h_t(\tau, \theta)] = E^{\theta_0} \left[ e^{ixX_{t_i}} \left( e^{ixX_{t_{i+1}}} - E^{\theta_0} \left[ e^{ixX_{t_{i+1}}} | X_{t_i} \right] \right) \right] = 0$$

As for the SVHJ model, $X_{t_i}$ is Markov, the latter holds true, so we have that indeed:

$$E^{\theta_0} [h_t(\tau, \theta_0)] = 0, \forall \tau \in \mathbb{R}^d, \pi\text{-almost everywhere}$$

Providing an analytic proof for the uniqueness of this result for $\theta_0$ is beyond the scope of this discussion given that we cannot derive an analytic characterization of the state-implying function.

**Assumption 5** - In order to prove the convergence of the criterion function we make use of an intermediate result from GW1988, Lemma 3.4, according to which, given the set-up of our Hilbert space norm given in Assumption 3, it is enough to prove that the moment function converges (uniformly). 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 of notations to match our set-up:

**Theorem 2.2** (cf. Politis and Romano 1994) Assume the $h_n(\tau, \theta)$ are stationary, bounded with probability one and have $\alpha$-mixing coefficients \(^{27}\) such that $\sum_{i=1}^{j} \alpha(i) \leq K j^r$, for all $1 \leq j \leq t_n$ where $r < 3/2$ and $K$ is the covariance operator defined in (4.11). Then $\sqrt{n} h_n(\tau, \theta_0)$ converges weakly to a Gaussian random element of $L^2(\pi)$ with a zero mean.

Given $h_n(\tau, \theta)$ is stationary (by Assumption 2) and bounded as proven, the only issue which remains is investigating the mixing properties of the moment condition function. As in Liebescher (2005), we argue that as $X_t$ is a stationary Markov process, under the additional assumption that it is also geometrically ergodic, the $\alpha$-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 (1993), page 367.

The uniformity in $\theta$ of the convergence, can be established by applying Lemma B.3 from \(^{27}\alpha(j) = \sup_{A,B} |P(A,B) - P(A)P(B)|, \text{where } A,B \text{ vary in the } \sigma \text{ algebras generated by } \{h_t, t \leq k\} \text{ and } \{h_t, t \geq j + k\}.$
Carrasco et al. (2007). Proving the stochastic equicontinuity required as a pre-condition in the latter lemma would require knowledge of the score of the state-implying function and of the score of the conditional characteristic function, neither of which are available in closed form for the SVHJ model.

Appendix C
Options Data-set: Descriptive Statistics

- Total sample span:

913 ‘Wednesdays’ 1 January 1996 31 August 2013 259691 options

- In-sample span:

723 ‘Wednesdays’ 1 January 1996 31 December 2009 145790 options

Descriptive statistics for the in-sample options\(^{28}\):

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call Price ($)</td>
<td>128.25</td>
<td>138.38</td>
<td>0.12</td>
<td>0.85</td>
<td>20.01</td>
<td>77.66</td>
<td>196.34</td>
<td>417.00</td>
<td>774.43</td>
</tr>
<tr>
<td>Implied Vol.</td>
<td>0.2376</td>
<td>0.0791</td>
<td>0.0814</td>
<td>0.1436</td>
<td>0.1800</td>
<td>0.2202</td>
<td>0.2773</td>
<td>0.3905</td>
<td>0.8618</td>
</tr>
<tr>
<td>Maturity</td>
<td>0.46</td>
<td>0.31</td>
<td>0.09</td>
<td>0.09</td>
<td>0.19</td>
<td>0.36</td>
<td>0.71</td>
<td>1.04</td>
<td>1.18</td>
</tr>
<tr>
<td>Moneyness (K/S)</td>
<td>0.96</td>
<td>0.18</td>
<td>0.50</td>
<td>0.64</td>
<td>0.84</td>
<td>0.97</td>
<td>1.07</td>
<td>1.27</td>
<td>1.49</td>
</tr>
</tbody>
</table>

- Out-of-sample span:

190 ‘Wednesdays’ 1 January 2010 31 August 2013 113901 options

Descriptive statistics for the out-of-sample options\(^{29}\):

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call Price ($)</td>
<td>138.63</td>
<td>146.17</td>
<td>0.125</td>
<td>0.55</td>
<td>15.75</td>
<td>85.91</td>
<td>224.02</td>
<td>444.19</td>
<td>645.61</td>
</tr>
<tr>
<td>Implied Vol.</td>
<td>0.1715</td>
<td>0.0721</td>
<td>0.0755</td>
<td>0.0958</td>
<td>0.1180</td>
<td>0.1520</td>
<td>0.2059</td>
<td>0.3038</td>
<td>0.7955</td>
</tr>
<tr>
<td>Maturity</td>
<td>0.44</td>
<td>0.31</td>
<td>0.09</td>
<td>0.09</td>
<td>0.17</td>
<td>0.34</td>
<td>0.72</td>
<td>1.02</td>
<td>1.18</td>
</tr>
<tr>
<td>Moneyness (K/S)</td>
<td>0.92</td>
<td>0.15</td>
<td>0.50</td>
<td>0.62</td>
<td>0.82</td>
<td>0.94</td>
<td>1.03</td>
<td>1.16</td>
<td>1.29</td>
</tr>
</tbody>
</table>

\(^{28}\)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 forward price as underlying.

\(^{29}\)Idem.
Appendix D
SVJ and SVVJ Model Specifications and Parameter Estimates

Following the same notations as in the SVHJ model set-up in equations (2.4)-(2.6) and (2.11)-(2.6), the stochastic volatility with Poisson jump component labeled SVJ and the stochastic volatility with volatility driven jump intensity labeled SVVJ models have the following specifications under the two equivalent probability measures:

SVJ

under $\mathbb{P}$:

\begin{align*}
    dy_t &= \left(\eta - \frac{1}{2}\right)v_t + (\mu - \mu^*)\lambda_t \ dt + \sqrt{v_t}dW_t^{(1),\mathbb{P}} + dJ_t^\mathbb{P} - \mu\lambda dt \\
    dv_t &= \kappa_v (\overline{v} - v_t) dt + \sigma_v \sqrt{v_t} \left(\rho dW_t^{(1),\mathbb{P}} + \sqrt{1 - \rho^2}dW_t^{(2),\mathbb{P}}\right). \tag{5.1}
\end{align*}

under $\mathbb{Q}$:

\begin{align*}
    dy_t &= \left(-\frac{1}{2}\right)v_t dt + \sqrt{v_t}dW_t^{(1),\mathbb{Q}} + dJ_t^\mathbb{Q} - \mu^*\lambda_t dt \\
    dv_t &= \kappa_v (\overline{v} - v_t) dt + \sigma_v \sqrt{v_t} \left(\rho dW_t^{(1),\mathbb{Q}} + \sqrt{1 - \rho^2}dW_t^{(2),\mathbb{Q}}\right). \tag{5.2}
\end{align*}

Table 3: SVJ estimated model parameters

<table>
<thead>
<tr>
<th>$\mu_j$</th>
<th>$\mu_j^{\mathbb{Q}}$</th>
<th>$\sigma_j$</th>
<th>$\eta$</th>
<th>$\kappa_v$</th>
<th>$\overline{v}$</th>
<th>$\sigma_v$</th>
<th>$\rho$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-13.21%</td>
<td>-18.77%</td>
<td>2.62%</td>
<td>2.47</td>
<td>5.13</td>
<td>0.009</td>
<td>0.195</td>
<td>-0.40</td>
<td>1.14</td>
</tr>
</tbody>
</table>

SVVJ

under $\mathbb{P}$:

\begin{align*}
    dy_t &= \left(\eta - \frac{1}{2}\right)v_t + (\mu - \mu^*)\lambda_1v_t \ dt + \sqrt{v_t}dW_t^{(1),\mathbb{P}} + dJ_t^\mathbb{P} - \mu_1\lambda v_t dt \\
    dv_t &= \kappa_v (\overline{v} - v_t) dt + \sigma_v \sqrt{v_t} \left(\rho dW_t^{(1),\mathbb{P}} + \sqrt{1 - \rho^2}dW_t^{(2),\mathbb{P}}\right). \tag{5.5}
\end{align*}

\begin{align*}
    dv_t &= \kappa_v (\overline{v} - v_t) dt + \sigma_v \sqrt{v_t} \left(\rho dW_t^{(1),\mathbb{P}} + \sqrt{1 - \rho^2}dW_t^{(2),\mathbb{P}}\right). \tag{5.6}
\end{align*}
under $Q$:

$$
\begin{align*}
dy_t &= \left(-\frac{1}{2}\right)v_t dt + \sqrt{v_t}dW_t^{(1),Q} + dJ_t^Q - \mu^*\lambda_1 v_t dt \\
dv_t &= \kappa_v (\bar{v} - v_t) dt + \sigma_v \sqrt{v_t} \left( \rho dW_t^{(1),Q} + \sqrt{1 - \rho^2}dW_t^{(2),Q} \right).
\end{align*}
$$

Table 4: SVVJ estimated model parameters

<table>
<thead>
<tr>
<th>$\mu_j$</th>
<th>$\mu_j^Q$</th>
<th>$\sigma_j$</th>
<th>$\eta$</th>
<th>$\kappa_v$</th>
<th>$\bar{v}$</th>
<th>$\sigma_v$</th>
<th>$\rho$</th>
<th>$\lambda_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.87%</td>
<td>-21.70%</td>
<td>3.70%</td>
<td>2.89</td>
<td>4.22</td>
<td>0.014</td>
<td>0.345</td>
<td>-0.45</td>
<td>28.12</td>
</tr>
</tbody>
</table>

References


