Modelling price and variance jump clustering using the marked Hawkes process

Journal of Financial Econometrics, 2023, 1–30
https://doi.org/10.1093/jjfinec/nbad007
Article

OXFORD

Modeling Price and Variance Jump Clustering Using the Marked Hawkes Process*

Jian Chen [1,2], **Michael P. Clements**[1], **and Andrew Urquhart**[1]

[1]ICMA Centre, Henley Business School, University of Reading, UK and [2]Department of Economics, University of Reading, UK

Address correspondence to Jian Chen, or e-mail: j.a.chen@reading.ac.uk.

Received January 31, 2022; revised February 21, 2023; editorial decision February 27, 2023; accepted March 2, 2023

Abstract

We examine the clustering behavior of price and variance jumps using high-frequency data, modeled as a marked Hawkes process (MHP) embedded in a bivariate jump-diffusion model with intraday periodic effects. We find that the jumps of both individual stocks and a broad index exhibit self-exciting behavior. The three dimensions of the model, namely positive price jumps, negative price jumps, and variance jumps, impact one another in an asymmetric fashion. We estimate model parameters using Bayesian inference by Markov Chain Monte Carlo, and find that the inclusion of the jump parameters improves the fit of the model. When we quantify the jump intensity and study the characteristics of jump clusters, we find that in high-frequency settings, jump clustering can last between 2.5 and 6 hours on average. We also find that the MHP generally outperforms other models in terms of reproducing two cluster-related characteristics found in the actual data.

Key words: jump clustering, marked Hawkes process, stochastic volatility, high-frequency data, Bayesian inference

JEL classification: C11, C52, C58

Understanding the behavior of large market movements or jumps in asset prices and its variance is essential to risk management. There are many studies that examine the importance of including both price and volatility jumps in asset pricing models (see, e.g., Eraker, 2004; Asgharian and Bengtsson, 2006; Barndorff-Nielsen and Shephard, 2006). These studies often assume serial independence of the jump components and use daily data in their empirical studies. However, less is known about the extent to which one jump presages subsequent jumps, especially at the intraday level. In this article, we study the clustering

* We thank the editor, the associate editor, and two anonymous referees for comments and suggestions that have greatly improved the manuscript.

behavior of jumps using intraday high-frequency data and a marked Hawkes process (MHP) embedded in a bivariate jump-diffusion model. We find evidence of self-excitation of jumps in both individual stocks and a broad equity index, and that jumps in prices and volatilities impact one another in an asymmetric manner. We also find evidence that the magnitudes of the jumps matter. Further, we demonstrate that the inclusion of jump parameters in the model significantly improves the fit of the model. Lastly, we simulate jump intensities under our models' settings and study cluster characteristics. We find the MHP model outperforms other benchmark models in reproducing some features of the actual data.

A potential explanation of why jumps might cluster comes from information asymmetry (Grossman, 1976). Suppose a publicly listed company's financial report indicates lower profitability. The most informed trader or investor would liquidate the company's stock from their portfolio, which may have a negative effect on the company's stock price. Less informed investors may follow suit, potentially producing negative variations, or jumps, clustering together. Lee (2012) also finds evidence of jumps in the stock market caused by information releases at both the macro level and firm level. These reasons suggest the potential importance of allowing for jump clustering in traditional asset pricing models, especially in recent times, given the higher-than-normal volatility in financial markets.

Some previous studies provide models of events clustering using self-exciting processes or Hawkes processes (HPs) (as proposed by Hawkes, 1971a,b). HP differs from a Poisson process, where events arrive randomly and independently of each other. HP relaxes the assumption of independent events arrivals and allows the underlying intensity of events to depend on past events. In addition, HP can be extended to MHP and multivariate MHP (MMHP) (e.g., Liniger, 2009). These models emphasize the impact of marked values (jump sizes) by allowing the intensities of events to depend on occurrences of past events. In addition, intensities are allowed to depend on the marked values attached to *other* past events (e.g., price jumps may depend on variance jumps).

There are many recent studies containing financial applications. Aït-Sahalia, Cacho-Diaz, and Laeven (2015) apply a multivariate HP to study jumps in international markets (see also Gresnigt, Kole, and Franses, 2016; Lee and Seo, 2022). Lee and Seo (2017) apply an MMHP to take both the impact of marked values and periodicity into account. They assume a symmetric structure in their study of price and variance. They find that intraday periodicity[1] is an important aspect of the modeling of price and variance dynamics (also see studies by Andersen and Bollerslev, 1997; Boudt, Croux, and Laurent, 2011). Lee and Seo (2017) cut off the first and last 30 minutes in every trading day to reduce the effects of intraday periodicity, and study the interaction between price and variance jumps. These studies provide a number of valuable insights regarding the application of Hawkes models. However, they use nonparametric methods to filter out financial market jumps. We are interested in incorporating the jump clustering feature into a general stochastic volatility model.

A strand of literature considers parametric models of price and variance dynamics for processes comprising both continuous and jump components. They accommodate Poisson

1 The variance of returns typically varies over a trading day, and tends to be highly correlated with trading volume, which is often higher during market opens and closes (Andersen and Bollerslev, 1997).

or Lévy processes in a continuous time semimartingale and commonly adopt the assumption that jumps arrive randomly, and increments are mutually independent (see e.g., Merton, 1976; Duffie, Pan, and Singleton, 2000; Eraker, 2004). More recently, potential interactions between price and variance jumps have been studied by Jacod et al. (2010) and Bandi and Renò (2016). Our goal is to investigate whether allowing for jump clustering provides superior models that better capture the characteristics of actual data.

Maneesoonthorn, Forbes, and Martin (2017) is similar to our study in some respects. They use high-frequency (intraday) data to detect jumps and calculate variance estimates and show that self-excitation is apparent in these estimates. However, their modeling uses daily-frequency data. Consequently, they may miss intraday clustering: a jump can raise the probability of a jump happening in the following few hours, but this may not show up in daily movements.

Our interest is investigating the stochastic volatility model in the high-frequency setting: a jump occurring in one day might, in fact, comprise a number of jumps within that day. At a daily level, any interdependencies in the intraday jumps would not be evident. Consider again a company disclosing negative news during trading hours. Traders can liquidate their assets quickly, perhaps within the same day, establishing intraday movements in prices and volatilities. So although jump clustering or self-excitation may be apparent in high-frequency data, it may not be so at a daily level. Other aspects may be of interest too: whether jump size is a determinant of clustering behavior, and the nature of the relationship between the magnitude of the jumps and the effect on the future intensity of jumps.

Hence our main contribution is to study the intraday dynamics of jump clustering within a general continuous-time asset pricing model. Our methodological contribution is to embed an MMHP in a price and variance state space model, and simultaneously estimate variances, jump magnitudes, and the effects price and variance jumps have on each other, etc. In addition, we consider intraday *periodicity*, and the potential this has to "hide" clustering behavior. In the estimation, we jointly estimate the static parameters and latent states by Bayesian Markov chain Monte Carlo (MCMC). We conduct a prior sensitivity analysis, and report diagnostic tests which attest to the reliability of the estimates.

In our empirical work, we find that both high-frequency stocks and an index exhibit self-exciting features. Our model also quantifies jump intensity. For example, using S&P 500 data, we show jump intensity or probability of jumps is lower than 1% during a "peaceful" period and becomes as high as 15% during a cluster. However, it decays quickly in the next few hours. We also show that price and variance jumps interact in an asymmetric fashion. Specifically, negative price jumps are more likely to be produced by past negative jumps rather than positive jumps, while positive jumps are as likely to be produced by past jumps of either sign. In addition, in some individual stocks, jump sizes are positively correlated with the following jump intensity: larger jumps tend to generate higher future jump intensity. Regarding the variance jumps, we find that variance jumps are more likely to produce negative return jumps rather than positive return jumps. However, the effect of variance jumps is much smaller than that of jumps in returns.

To judge the support the data lends to the various models in our study, we report the Deviance Information Criteria (DIC) and the Bayes Factor. These are shown to favor the modeling of returns and variance with MMHP models. We also evaluate the range of

models we consider in terms of their ability to generate features found in the data—namely aspects of the clustering of jumps. This goes beyond the general fit of the models to the data, and highlights the ability of the models to capture aspects of particular importance. The MMHP generally outperforms the other models in terms of reproducing these characteristics. There is some evidence that the simpler HP is already sufficient to reproduce some features of the clustering of variance jumps. The findings in this article may be relevant to risk managers, high-frequency traders, and other practitioners, who may benefit from the knowledge of the interdependencies of jumps.

The article is organized as follows. Section 1 contains the theoretical framework. Section 2 presents the estimation method, along with jump detection methods. Section 3 discusses the empirical findings. Section 4 describes the simulation tests. We conclude in Section 5. Some technical results are confined to an Appendix.

1 Theoretical Setup

In this section, we introduce a continuous-time price and variance jump-diffusion process with an MMHP embedded, and present the discretized form of the model.

1.1 Continuous-Time Representation of Price and Variance Process

We let P_t be the natural logarithm of asset prices and V_t be the variance at time t, and consider the following jump-diffusion process:

$$dP_t \cdot S_t = \mu dt + \sqrt{V_t} dW_t^P + \xi_t^{P+} dN_t^{P+} + \xi_t^{P-} dN_t^{P-} \tag{1}$$

$$dV_t = \gamma(\theta - V_t)dt + \sigma_V \sqrt{V_t} dW_t^V + \xi_t^V dN_t^V, \tag{2}$$

where S_t is a periodic component, μ is a drift term, γ and θ denote the mean reversion speed and the long-run variance mean, respectively, and σ_V refers to the volatility of volatility. W_t^P and W_t^V are Wiener processes of return and variance, respectively, and we let increments of them be correlated $E(dW_t^P, dW_t^V) = \rho dt$. In terms of the jump components in the processes, we separate price jumps into positive and negative groups. However, we only consider positive variance jumps in our study, with decreases of variance captured by γ and θ. $\{\xi_t^{P+}, \xi_t^{P-}, \xi_t^V\}$ denote the size of the jumps. For ease of estimation, we let the sizes of the price jumps jointly follow a normal distribution $\xi_t^{P+}, \xi_t^{P-} \sim N(\mu_P, \sigma_P)$, and those of variance jumps follow an exponential distribution with mean μ_V, $\xi_t^V \sim \exp(\mu_V)$. In terms of jump components, we employ the following three-dimensional MHP to present the jump components in the processes:

$$P(dN_t^i = 1) = \lambda_t^i dt, \quad i = \{P+, P-, V\}, \tag{3}$$

where λ_t^i denotes the intensity of the counting process, which is defined as an MMHP in Equation (4). This constitutes a departure from the literature. The continuous-time asset pricing model typically assumes a constant jump intensity (see, e.g., Duffie, Pan, and Singleton, 2000; Eraker, 2004; Bandi and Renò, 2016). Our model allows mutually dependent jump intensities, which may be correlated with jump size. Another departure from the literature is our use of intraday data to investigate the dynamics of price, variance, and jump clustering, as opposed to the use of daily data in most previous studies.

1.2 A Three-Dimensional MHP

Here we introduce an MMHP. The corresponding intensity of the jump processes in Equation (3) is given by:

$$\lambda_t^i = \lambda_0^i + \sum_{q=1}^{d} \vartheta_q^i \int_{[0,t)\times\Xi} \phi_q^i(t-s)\omega_q^i(\xi_q) N_q(ds \times d\xi), \tag{4}$$

where λ_0^i denotes the immigration intensity of dimension i (three dimensions are indicated in Equation 3) which is constant; $\boldsymbol{\vartheta}$ is a branching coefficient matrix; ϕ_q^i is a decay function such that $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ and we assume an exponentially decaying kernel $\phi_q^i(t-s) = \beta_q^i e^{-\beta_q^i(t-s)}, \beta_q^i > 0$. $\omega_q^i(\xi_q)$ is an impact function of jump sizes, such that $\omega : \mathbb{R} \to \mathbb{R}_+$.

Remark 1 (Interpretation) *Under a general HP, the immigrants arrive as a Poisson process with immigration intensity λ_0^i and jump sizes ξ. They generate subsequent arrivals (further jumps), with an intensity which decays exponentially with the speed parameter β. $\boldsymbol{\vartheta}$ is defined as a (3×3) branching matrix $\boldsymbol{\vartheta} := \vartheta_q^i; q, i \in \{P+, P-, V\}$ and ϑ_q^i measures the expected number of jumps in dimension i that is produced by jumps in dimension q. The impact function $\omega_q^i(\xi_q)$ measures the impact of the jump sizes ξ_q in the dimension q. Further details of MMHP and its interpretations are provided by Liniger (2009).*

Remark 2 (A simple example) *For a simple example of MMHP, given an occurrence of a jump in dimension q (size: ξ_q), the underlying intensity of dimension i will increase from λ_0^i to $\lambda_0^i + \vartheta_q^i \cdot \omega_q^i(\xi_q) \cdot \beta_q^i$. Then, the incremental part of the intensity $\vartheta_q^i \cdot \omega_q^i(\xi_q) \cdot \beta_q^i$ will decay at a speed of $\phi_q^i(dt) = \beta_q^i e^{-\beta_q^i \cdot dt}$ for every dt.*

Further details, including stationary assumptions, are in Daley and Vere-Jones (2003), with which most of our mathematical setup is in line.

Remark 3 (Impact Functions) *Inspired by Liniger (2009), we normalized our impact functions to satisfy some stationary conditions, denoting $\tilde{\omega}_q(\cdot)$ as the impact function before being normalized. The normalized impact function is given by:*

$$\omega_q^i(\xi_q) = \frac{\tilde{\omega}_q^i(\xi_q)}{\mathbb{E}[\tilde{\omega}_q^i(\boldsymbol{\xi})]} \tag{5}$$

and we consider four impact functions (before normalization) as follows:

$$\begin{aligned}
\tilde{\omega}_q^i(\xi_q) &= \tilde{\alpha}_q^i + \tilde{\beta}_q^i|\xi_q| + \tilde{\gamma}_q^i|\xi_q|^2 \quad (I)\\
\tilde{\omega}_q^i(\xi_q) &= |\xi_q|^{\tilde{\alpha}_q^i} \quad (II)\\
\tilde{\omega}_q^i(\xi_q) &= e^{\tilde{\alpha}_q^i|\xi_q|} \quad (III)\\
\tilde{\omega}_q^i(\xi_q) &= \tilde{\alpha}_q^i + \tilde{\beta}_q^i \log(1+|\xi_q|) \quad (IV)
\end{aligned} \tag{6}$$

We assume the parameters in the above impact functions $\{\alpha, \beta, \gamma\}$ satisfy those within each function, as at least one of the parameters is strictly positive. Also, noticeably, the jump sizes have no impact on intensity processes when $\omega(\xi) \equiv 1$. In addition, we assume an identical independent distribution of jump sizes.

1.3 Discretized Form of Return and Variance Processes

Similar to previous studies, we apply an Euler discretization to the processes with $\Delta t = \frac{1}{79\times 252}$ (equivalent to five minutes) and obtain the following forms:

$$(P_t - P_{t-\Delta t})S_t = \mu + \sqrt{V_{t-\Delta t}}\epsilon_t^P + \xi_t^{P+}\Delta J_t^{P+} + \xi_t^{P-}\Delta J_t^{P-} \tag{7}$$

$$V_t = \alpha_v + (1+\beta_v)V_{t-\Delta t} + \sigma_V\sqrt{V_{t-\Delta t}}\epsilon_t^V + \xi_t^V\Delta J_t^V, \tag{8}$$

where P_t and V_t denote the logarithm of the asset price and the variance at time t, S_t is a periodic component and $S_t = I_{t,\tau}\cdot f_\tau$, where $\tau = 1, 2, \ldots, 79$, and $I_{t,\tau}$ is an indicator variable such that $I_{t,\tau} = 1$ when τ corresponds to time t.[2] μ denotes the drift term, and σ_V denotes the volatility of volatility. The mean reversion speed and level of variance are translated to α_v, and β_v, where $\alpha_v = \gamma\theta$ and $\beta_v = -\gamma$ (γ, θ are mean reversion speed and level of variance, see Section 1.3). ϵ_t^P and ϵ_t^V are two random variables that follow a normal distribution $N(0,1)$ with correlation $corr(\epsilon_t^P, \epsilon_t^V) = \rho$. $\xi_t^i, i = \{P+, P-, V\}$ denote jump magnitudes, which we specify by:

$$\xi_t^{P+}, \xi_t^{P-} \sim N(\mu_P, \sigma_P) \tag{9}$$

$$\xi_t^V \sim \exp(\mu_V). \tag{10}$$

The sizes of price jumps jointly follow a normal distribution, and variance jump sizes follow an exponential distribution with a parameter μ_V, while $\Delta J_t^i = J_t^i - J_{t-\Delta t}^i, i = \{P+, P-, V\}$ is a Bernoulli random variable with corresponding time-varying intensity λ_t^i, so:

$$\Delta J_t^i \sim Bernoulli(\lambda_t^i), \quad i = \{P+, P-, V\}, \tag{11}$$

$\Delta J_t^i = 1$ can be viewed as an occurrence of a jump, and λ_t^i can be regarded as the probability of a jump happening at time t. We consider a number of cases. The simplest and most common is that jumps follow a Poisson process, and the underlying intensity λ_t^i is constant at $\lambda_t^i \equiv \lambda_0^i$, such that the probability of a jump happening is the same over time and jumps arrive independently. For a multivariate HP, without considering the impact of marks, the intensity process is specified as follows:

$$\lambda^i(t) = \lambda_0^i + \sum_q \vartheta_q^i \sum_{0<s<t} \phi_q^i(t-s), \quad i, q = \{P+, P-, V\}, \tag{12}$$

where ϑ_q^i is a branching matrix and $\phi_q^i(\cdot)$ is an exponentially decaying kernel (see details of interpretations in Remarks 1 and 2). In addition, by adding an impact function on jump sizes $\omega_q^i(\cdot)$, an MMHP can be introduced to define the intensity of dimension i at time t as follows:

$$\lambda^i(t) = \lambda_0^i + \sum_q \vartheta_q^i \sum_{0<s<t} \phi^i(t-s)\omega_q^i(\xi_q), \quad i, q = \{P+, P-, V\}, \tag{13}$$

where $\omega_q^i(\cdot)$ is the impact function. When $\omega_q^i(\cdot) \equiv 1$, we obtain the previous multivariate HP model in which jump size has no effect on intensity. We consider four different impact functions (see Remark 3). By adding components of the right-hand side of Equation (13) in

2 We employ five-minute data, and, therefore, have seventy-nine observations in a trading day. The time variable t is indexing every five minutes, and when τ corresponds to time t, we adjust returns by the periodic component f_τ.

addition to λ_0^i, intensity, or probability of a jump happening is allowed to depend on past jumps through the branching matrix ϑ_q^i and jump sizes through impact functions $\omega_q^i(\cdot)$.

Duffie, Pan, and Singleton (2000) proposed a stochastic volatility with correlated jumps (SVCJ) model. They specify a dependent structure between the size of the price and the variance jump. But price and variance jump in the model are still assumed to arrive independently with constant underlying intensities. In our model, we assume the underlying intensities of jumps are time-varying and consider how jump size affects the underlying intensities.

2 Bayesian Estimation

We follow Eraker (2004) in using a Bayesian approach to estimate the stochastic volatility model. Estimating parameters in the HP can be easily embedded in the estimation of the price and variance dynamics. The Bayesian analysis of HP is provided by Rasmussen (2013). We denote the static parameter vector as $\Theta = \{\mu, \alpha_v, \beta_v, \sigma_V, \rho, \lambda_0^i, \beta^i, \Phi^i\}$, where Φ^i denotes parameters in the impact function (6) and depends on which function is used. We also denote the latent state vector as $\Omega_t = \{\Delta J_t^{P+}, \Delta J_t^{P-}, \Delta J_t^V, \xi^{P+}, \xi^{P-}, \xi^V, \boldsymbol{\vartheta}, V_t\}$. Then, the joint posterior distribution using the Bayes formula is:

$$p(\Theta, \Omega_t | P_t) \propto p(P_t | \Theta, \Omega_t) p(\Omega_t | \Theta) p(\Theta) \tag{14}$$

The posterior given in Equation (14) is apparently not available in closed form due to the complexity of the processes. Therefore, we adopt a MCMC method to generate a sequence of draws on parameters and latent variables and simulate the posterior. The specification of priors and details of the algorithm are in Appendix A.

However, the estimation is subject to a significant computational burden. Therefore, we make an assumption regarding Equation (12), namely that:

$$\lambda^i(t) = \lambda_0^i + \sum_q \vartheta_q^i \sum_{\iota < s < t} \phi_q^i(t-s)\omega_q^i(\xi_q), \quad i, q = \{P+, P-, V\}, \tag{15}$$

where $\iota = -log\left(\frac{c}{\beta_q^i}\right)$ and we take $c = 0.01\%$. Note in the original Equation (12), $\iota = 0$. This assumption on ι avoids the need to consider the impact of all the past jumps. Instead, we consider only those jumps whose impact has not decayed by over 99.99%. In other words, we assume that when the incremental intensity raised by a past jump has decayed by over 99.99%, it can be ignored. This greatly facilitates the calculation of the marginal likelihood of the jump processes, and speeds up the MCMC estimation algorithm.

3 Empirical Application

In this section, we introduce our dataset and set out the range of models to be considered. We first report the parameter estimates of the price and variance process and MMHP kernels. We then compare the models in terms of their fit to the data, and finally present a prior sensitivity analysis and some diagnostic tests regarding the convergence of the MCMC. We estimate the model with 40,000 loops, with the first 10,000 as the burn-in period.

3.1 Data

We retrieved five-minute price data from Bloomberg for four individual stocks across different industries; namely, Apple (AAPL), Boeing (BA), J.P. Morgan (JPM), Coca-Cola (KO), and the S&P 500 Index from January 3, 2012 to December 31, 2019. We chose these stocks as they are large firms that have high liquidity, are from different industries, and are likely to be followed by analysts and therefore react to news quickly. We also examine three ETF data, namely iShares FTSE China Index Fund, iShares MSCI Spain ETF, and SPDR S&P 500 Trust ETF. We cleaned the data following the approach in Barndorff-Nielsen et al. (2008). We set the time unit $\Delta t = \frac{1}{79 \cdot 252}$. We do not discard overnight returns, but instead capture the large variation of overnight returns by an intraday periodicity term in some instances. In Figure 1, we plot the log return ($Y_t = P_t - P_{t-\Delta t}$), the log return minus estimated jumps ($Y_t - \xi_t^P$), and estimated volatility ($\sqrt{V_t}$) of the individual stocks and the S&P 500 Index, for the period 2012–2019.

3.2 Parameters Estimated in Price and Variance Process

We estimate Equations (7) and (8), but with different specifications for the intensities of jumps $\{J_t^{P+}, J_t^{P-}, J_t^V\}$. We specify seven models ($\mathcal{M}_1$ to $\mathcal{M}_7$) in Table 1 representing alternative ways of modeling jump intensity. $\mathcal{M}_1$ is our benchmark model, corresponding to the SVCJ model of Duffie, Pan, and Singleton (2000).[3] $\mathcal{M}_2$ and $\mathcal{M}_3$ exhibit jump clustering, modeled by the multivariate HP, but $\mathcal{M}_2$ ignores the intraday periodicity and assumes $S_t \equiv 1$. $\mathcal{M}_4$ to $\mathcal{M}_7$ model jumps by MMHP with different impact functions.

In Table 2, we report the parameter estimates of the price and variance processes for $\mathcal{M}_1$. We do not record the estimates of these parameters for the other model specifications, because they change little, and are of secondary interest compared to the jump parameters.

3.3 Jump Clustering Parameters

In this subsection, we report the parameter estimates of the Hawkes kernels in $\mathcal{M}_2$ and $\mathcal{M}_3$. We set the impact function to $\omega_q(\xi_q) \equiv 1$, which implies jump sizes do not impact future jump intensities. A summary of the results is in Table 3, with full results reported in the Supplementary Appendix. In Table 3, we give the posterior means of the parameters, and their significance.[4] For price jumps, there is a clear finding that, without periodicity ($\mathcal{M}_2$), $\{\vartheta_{p+}^{p+}, \vartheta_{p-}^{p+}, \vartheta_{p+}^{p-}, \vartheta_{p-}^{p-}\}$, are significantly greater than 0. Although this self-excitation feature appears in individual stocks data, it does not do so for the S&P 500 Index. This result is in tune with previous studies, such as Foschi, Lilla, and Mancini (2019). However, with periodic components ($\mathcal{M}_3$), the price jumps in the S&P 500 also display self-excitation. We find that the decay speed of some variance jumps is faster than others judged by the values of $\{\beta_v^v, \beta_{p-}^v, \beta_{p+}^v\}$.

3 They assume a constant jump intensity but allow jump sizes be correlated: $\xi_t^P \sim \mathcal{N}(\mu_J + \rho_J \xi_t^V, \sigma_J)$.

4 To give a broad view of whether estimated parameters are different from 0, we look at whether 95% of their posterior distribution is greater than 0. For those posteriors that are truncated at 0 (they are not allowed to be sampled at any negative values), we look at whether their posterior means minus two times their posterior standard deviations is greater than 0. We refer to this as "significance" throughout this article.

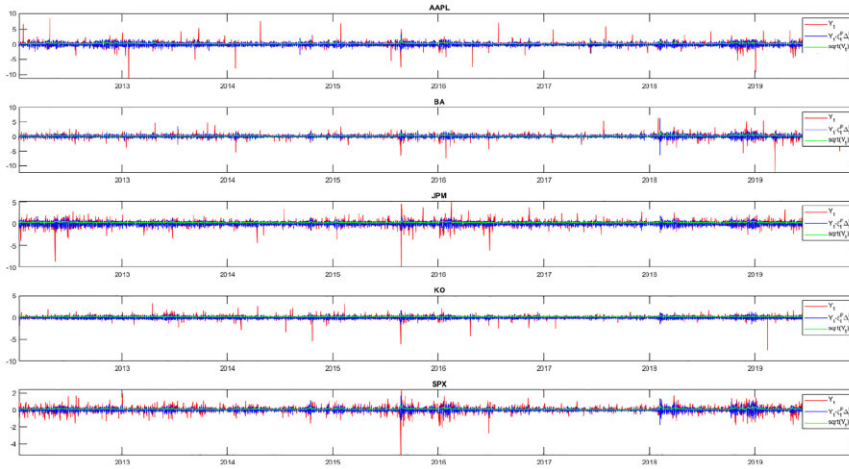

Figure 1 Log return, estimated volatility and return minus return jumps of individual stocks and S&P 500 Index.

Notes: This figure presents log returns ($Y_t = P_t - P_{t-1}$), return minus estimated return jumps ($Y_t - \xi_t^P$), and the estimated volatility ($\sqrt{V_t}$).

Table 1 Specification of $\mathcal{M}_1$ to $\mathcal{M}_7$

	Price and variance process with jumps modeled by
$\mathcal{M}_1$	Correlated jump size components (SVCJ) model
$\mathcal{M}_2$	HP without intraday periodicity ($S_t \equiv 1$)
$\mathcal{M}_3$	HP with intraday periodicity
$\mathcal{M}_4$	MHP with impact function (6, *I*)
$\mathcal{M}_5$	MHP with impact function (6, *II*)
$\mathcal{M}_6$	MHP with impact function (6, *III*)
$\mathcal{M}_7$	MHP with impact function (6, *IV*)

The results also suggest asymmetry of the branching coefficient matrix $\boldsymbol{\vartheta}$. Specifically, posterior mean of ϑ_{p-}^{p+} is not statistically significantly different from that of ϑ_{p+}^{p+} (0.046 and 0.044, first column in $\mathcal{M}_3$ panel). There is no evidence that positive jumps are more or less likely to be produced by positive as opposed to negative price jumps. However, a posterior mean of ϑ_{p-}^{p-} is statistically significantly greater than ϑ_{p+}^{p-} (0.046 and 0.088, 2nd column in $\mathcal{M}_3$ panel). This suggests negative jumps are more likely to follow from negative jumps than to be produced by positive jumps. There are some other studies that look at the heterogeneity of positive and negative return variations. For example, Bollerslev, Li, and Zhao (2020) find that positive and negative return variations are priced differently in asset prices.

Furthermore, we find that variance jumps are more likely to produce negative return jumps than positive ones (0.02 and 0.025 in Column 3, $\mathcal{M}_3$; 0.038 and 0.048 in the last column, $\mathcal{M}_3$). This is consistent with the results in Ait-Sahalia, Fan, and Li (2013), which

Table 2 Posterior mean and standard deviation of parameters in price and variance process

	AAPL		BA		JPM		KO		SPX	
	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.
$\mu(*10^{-4})$	3.822	3.721	0.602	3.533	1.128	3.018	2.980	2.304	**6.610**	1.815
$\alpha_v(*10^{-3})$	**1.461**	0.057	**1.572**	0.044	**0.924**	0.032	**0.610**	0.021	**0.171**	0.005
β_v	**−0.017**	0.001	**−0.024**	0.001	**−0.018**	0.001	**−0.028**	0.001	**−0.014**	0.001
$\sigma_V^2(*10^{-3})$	**2.039**	0.057	**1.984**	0.046	**1.097**	0.031	**0.634**	0.019	**0.272**	0.008
ρ	**−0.161**	0.010	**−0.070**	0.009	**−0.129**	0.011	**−0.053**	0.010	**−0.237**	0.010
μ_P	−0.019	0.031	−0.038	0.031	0.013	0.030	−0.005	0.023	−0.039	0.015
σ_P	**0.7264**	0.0566	**0.7805**	0.1239	**0.4932**	0.0498	**0.2650**	0.0272	**0.1586**	0.0172
μ_V	**0.0025**	0.0005	**0.0021**	0.0004	**0.0040**	0.0011	**0.0050**	0.0016	**0.0054**	0.0018

Notes: Parameters are estimated with data inflated by 79 × 100 times. Significant values (95% of the posterior distribution is greater than 0) are marked in bold.

Table 3 Summary of select parameter posteriors

	Individual stocks			S&P 500			ETF		
	$p+$	$p-$	v	$p+$	$p-$	v	$p+$	$p-$	v
				ϑ in $\mathcal{M}_2$					
$p+$	0.02 (0/4)	0.014 (0/4)	0.015 (0/4)	0.007 (0/1)	0.006 (0/1)	0.015 (0/1)	0.016 (0/3)	0.019 (0/3)	0.016 (0/3)
$p-$	0.043 (4/4)	0.036 (4/4)	0.015 (0/4)	0.02 (0/1)	0.016 (0/1)	0.012 (0/1)	0.029 (2/3)	0.047 (2/3)	0.010 (1/3)
v	0.024 (0/4)	0.017 (0/4)	0.009 (0/4)	0.014 (0/1)	0.011 (0/1)	0.027 (0/1)	0.003 (0/3)	0.008 (0/3)	0.037 (0/3)
				ϑ in $\mathcal{M}_3$					
$p+$	0.046 (4/4)	0.046 (4/4)	0.02 (2/4)	0.039 (1/1)	0.034 (1/1)	0.023 (1/1)	0.051 (3/3)	0.031 (3/3)	0.038 (2/3)
$p-$	0.044 (4/4)	0.088 (4/4)	0.025 (4/4)	0.052 (1/1)	0.077 (1/1)	0.033 (0/1)	0.075 (3/3)	0.091 (3/3)	0.048 (3/3)
v	0.017 (0/4)	0.009 (0/4)	0.026 (2/4)	0.011 (0/1)	0.009 (0/1)	0.013 (0/1)	0.012(0/3)	0.102(2/3)	0.116 (2/3)
				$\tilde{\beta}$in $\mathcal{M}_4$					
$p+$	0.35 (3/4)	0.35 (2/4)	0.325 (0/4)	0.351 (0/1)	0.452 (1/1)	0.309 (0/1)	0.382 (2/3)	0.015 (0/3)	0.057 (0/3)
$p-$	0.3 (1/4)	0.375 (4/4)	0.3 (0/4)	0.37 (0/1)	0.382 (1/1)	0.358 (0/1)	0.252 (1/3)	0.651 (3/3)	0.267 (0/3)
$\boldsymbol{v}$	10.775 (0/4)	14.425 (1/4)	15.5 (3/4)	12.225 (0/1)	16.155 (0/1)	8.217 (0/1)	12.4 (1/3)	19.5 (1/3)	10.1 (1/3)
				$\tilde{\gamma}$ in $\mathcal{M}_4$					
$p+$	0.025 (0/4)	0.05 (0/4)	0.05 (0/4)	0.053 (1/1)	0.021 (0/1)	0.085 (0/1)	0.042 (1/3)	0.028 (0/3)	0.036 (0/3)
$p-$	0.05 (0/4)	0 (0/4)	0.075 (0/4)	0.03 (0/1)	0.024 (0/1)	0.043 (0/1)	0.012 (0/3)	0.039 (1/3)	0.024 (0/3)
v	11.65 (3/4)	12.6 (3/4)	8.1 (2/4)	28.024 (1/1)	29.456 (1/1)	52.929 (1/1)	27.56 (1/3)	19.6 (2/3)	94.9 (3/3)
				$\tilde{\beta}$ in $\mathcal{M}_7$					
$p+$	0.825 (4/4)	0.85 (3/4)	0.875 (0/4)	0.811 (1/1)	0.739 (0/1)	0.795 (0/1)	1.157(2/3)	1.542 (2/3)	0.108 (0/3)
$p-$	0.775 (4/4)	0.75 (4/4)	0.775 (1/4)	0.78 (1/1)	0.782 (1/1)	0.781 (0/1)	1.368 (3/3)	2.015 (3/3)	0.095 (0/3)
v	13.3 (2/4)	14.925 (2/4)	15.175 (1/4)	10.245 (0/1)	18.212 (0/1)	14.373 (0/1)	9.62(1/3)	12.548 (2/3)	9.541 (1/3)

Notes: This table presents posterior means and the number of significant results of some parameters in models. The rows record the responses to the dimensions of variables in the columns. For example, ϑ denotes the mean number of jumps of one dimension that is produced by another dimension. So, the 0.043 (ϑ_{p-}^{p+} in the 2nd row, 1st column) means in individual stock data, there are, on average, 0.043 negative price jumps ($p-$) produced by positive price jumps ($p+$). In addition, there are in total four individual stocks and this parameter is significant for all of them, hence, (4/4).

find jumps to be a source of the leverage effect. In addition, the impact of variance jumps on other types of jumps is smaller, and often associated with a faster rate of decay.

3.4 Impact of Jump Sizes

We next examine the impacts of jump sizes. The results of the jump-size-related parameters, $\tilde{\beta}$ and $\tilde{\gamma}$ in $\mathcal{M}_4$, and $\tilde{\beta}$ in $\mathcal{M}_7$, are also reported in Table 3. These parameters are from the impact functions (see Equation 6) of MMHPs.

In Table 3, some of the coefficients of jump size are significantly greater than 0, and are associated with a high decay rate (β). For example, given $\beta = 0.5$, the intensity heightened by past jumps will decay to 5% of the original after five time-units (30 minutes), and it will decay to 0.3% within the same time if $\beta = 1$. However, the percentage will be 41% if $\beta = 0.15$. Therefore, the magnitude of β, and not just its statistical significance, plays key a role in determining the likelihood of future jumps being generated.

Table 3 also shows that some of the coefficients $\{\tilde{\beta}^{p+}_{p+}, \tilde{\beta}^{p-}_{p-}\}$, $\{\tilde{\gamma}^{p+}_{v}, \tilde{\gamma}^{p-}_{v}, \tilde{\gamma}^{v}_{v}\}$ in $\mathcal{M}_4$ and $\{\tilde{\beta}^{p+}_{p+}, \tilde{\beta}^{p+}_{p-}, \tilde{\beta}^{p-}_{p+}, \tilde{\beta}^{p-}_{p-}\}$ in $\mathcal{M}_7$ are statistically significantly positive. This suggests that the underlying intensities of jumps may be affected by both positive and negative jumps, and that larger price jumps may generate higher intensities for future price jumps.

Moreover, the interaction between price jumps and variance jumps is captured by the coefficients of squared jump sizes $\{\tilde{\gamma}^{v}_{p+}, \tilde{\gamma}^{v}_{p-}\}$, but they are less significant in the log transform of jump sizes.

The parameter estimates for $\mathcal{M}_5$ and $\mathcal{M}_6$ suggest strong evidence of self- and cross-exciting behavior. The parameters in $\mathcal{M}_5$ are mostly significant, and $\{\tilde{\alpha}^{p+}_{p+}, \tilde{\alpha}^{p+}_{p-}, \tilde{\alpha}^{p-}_{p+}, \tilde{\alpha}^{p-}_{p-}, \tilde{\alpha}^{v}_{v}\}$ in $\mathcal{M}_6$ are all significant.

Finally, to better bring out some of the features of the models, we provide a graphical analysis of some of the latent variables of our model ($\mathcal{M}_7$ in Table 1), estimated using S&P 500 Index data from December 12, 2018 to December 26, 2018 (two weeks, ten trading days) in Figure 2. Note the flat line of price and return on December 24, 2018 is due to the early close (2:00 P.M.) of the stock market on Christmas eve. The intraday periodicity component (third figure) is estimated as having a clear U-shape. The last three figures plot the estimated return jumps, volatility jumps, and the negative jump intensity. We collect these jumps and jump intensities from the last iteration of the MCMC algorithm. We see that on December 21, 2018, the market was turbulent, and the index dropped by 3.1% at the end of the day. The model estimated volatility (green line) shows a clear increase during that day, but this is insufficient to fully explain the variability of returns. Instead, three negative jumps are identified, creating a cluster of negative return jumps. This cluster lasted 1.5 hours in total, and the underlying intensity of negative jumps is as high as 15%. The index dropped by 1.6% during the cluster. Although there are four spikes in the estimated intensity, only three negative jumps are shown, with the additional spike being created by cross-excitation from a positive return jump. We provide two more examples of return and variance jump clustering in the Supplementary Appendix.

3.5 Model Fit

To assess the goodness-of-fit of our models ($\mathcal{M}_1$ to $\mathcal{M}_7$), we calculate the DIC of Spiegelhalter et al. (2002) and the Bayes factor of Kass and Raftery (1995). The DIC is calculated from the log-likelihood penalized by model complexity, and can be easily obtained

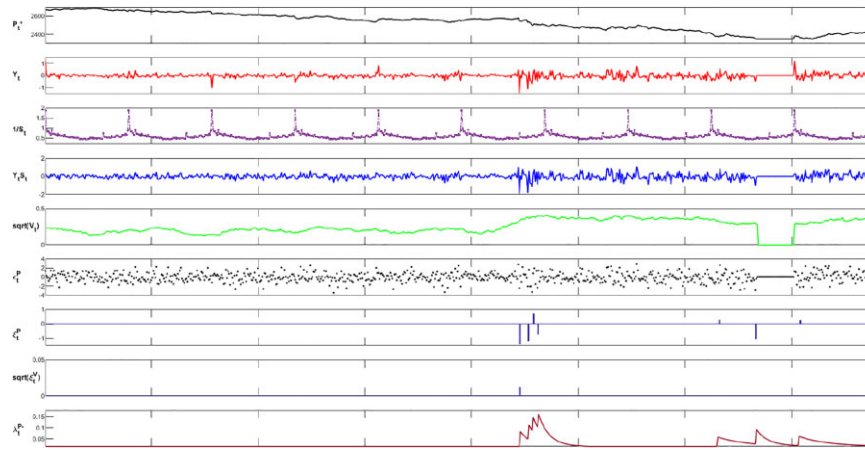

Figure 2 Latent variables of S&P 500 Index two weeks from December 12, 2018 estimated in $\mathcal{M}_7$.

Notes: This figure plot some latent variables of the model ($\mathcal{M}_7$) estimated using S&P 500 Index data from December 12, 2018 to December 26, 2018 (two weeks, 10 trading days). The first figure (P_t^*) is the price of S&P 500, and Y_t denotes the log return. We collect the last three variables (return jumps, volatility jumps, and the negative return jump intensity) from the last iteration of the MCMC algorithm. Note the flat line on December 24, 2018 is due to the early close (2:00 P.M.) of the stock market on Christmas eve. Note on December 21, 2018, the market entered a turbulent trading period and the index dropped by 3.1% at the close. Three negative jumps are identified, which created a cluster of negative return jumps and lasted 1.5 hours. The index dropped by 1.6% during the cluster.

in an MCMC algorithm. A lower value indicates a better fit of the model. Table 4 shows that $\mathcal{M}_1$ (modeling jumps as a Poisson process) has the highest DIC, while the MMHP with impact function (6, *VII*) is rewarded with the lowest DIC ($\mathcal{M}_7$). In addition, three of the four MMHP models ($\mathcal{M}_4$, $\mathcal{M}_5$, $\mathcal{M}_7$) outperform the multivariate-HP model $\mathcal{M}_3$ which does not feature an impact from jump size. Finally, not surprisingly, $\mathcal{M}_6$ with an exponential impact function underperforms $\mathcal{M}_3$, since the parameters in $\mathcal{M}_6$ suggest there are few self-exciting features in the data.

Table 5 reports the log values of the Bayes factor across the seven models. We find that $\mathcal{M}_1$ is inferior to all the other models. In addition, the results show a large discrepancy between other models against $\mathcal{M}_2$ and $\mathcal{M}_3$, which suggests the importance of considering intraday periodicity. In addition, the log Bayes factor of $\mathcal{M}_6$ against $\mathcal{M}_3$ is negative, which is in line with the results for the DIC.

It is unclear whether $\mathcal{M}_4$ or $\mathcal{M}_7$ provides the better fit. Overall, $\mathcal{M}_7$ presents better results, but $\mathcal{M}_4$ prevails for the BA data. The marginal likelihood estimates for the different models can be seen in Appendix B.

3.6 Robustness Check

Since our model is heavily parameterized, it is important to ensure the estimation is reliable. We perform a prior sensitivity analysis and some diagnostic tests on the convergence of the MCMC.

Table 4 DIC of $\mathcal{M}_1$ to $\mathcal{M}_7$

	AAPL	BA	JPM	KO	SPX
$\mathcal{M}_1$	276,836	303,607	296,915	394,377	510,339
$\mathcal{M}_2$	274,364	301,632	295,344	393,650	506,624
$\mathcal{M}_3$	272,104	298,750	291,587	391,447	500,386
$\mathcal{M}_4$	267,014	247,398	273,056	383,631	491,936
$\mathcal{M}_5$	271,907	294,875	289,265	389,559	499,173
$\mathcal{M}_6$	272,454	298,988	291,770	391,608	500,323
$\mathcal{M}_7$	264,881	249,458	267,975	382,774	490,330

3.6.1 Prior sensitivity analysis

We conduct a sensitivity analysis of the prior to investigate if the estimates of the latent variables and some of the key variables of interest are unduly sensitive to the values chosen for the priors. We follow Chib, Nardari, and Shephard (2002) and Nakajima and Omori (2009), who specify alternative priors of variance and jump-related parameters to test their sensitivities. Our selected values are summarized in Table 6. We examine the prior of the mean-reversion speed parameter β_v and the jump variance parameter σ_P^2. We also consider the jump-clustering-related parameters $\{\boldsymbol{\vartheta}, \boldsymbol{\beta}\}$. The alternative settings of jump-clustering-related parameters are inspired by Rasmussen (2013). In the tests, we focus on $\mathcal{M}_4$ and $\mathcal{M}_7$, because these models contain jump clustering parameters and are the best-performing models in terms of fit.

The results of the sensitivity analysis are presented in Table 7. Overall, parameters in the price and variance processes are not overly affected by different prior settings. Some parameters $\{\mu, \rho, \sigma_Y\}$ vary more than others $(\{\alpha_v, \beta_v, \sigma_V^2, \mu_V\})$ under alternative settings. Of the jump clustering parameters, Pr_1 has the biggest influence in $\mathcal{M}_7$. Nevertheless, the deviations of the posterior means from their original values (Pr_0) are no higher than 7.5%. Posterior standard deviations are more affected. The full table of results for the sensitivity analysis is provided in the Supplementary Appendix.

3.6.2 MCMC Convergence diagnostic tests

We perform a Geweke diagnostic test (Geweke et al., 1991) and a Raftery–Lewis test (Raftery and Lewis, 1991) to examine the convergence of the parameter estimates.

The Geweke test divides a Markov chain of a sampled variable into two parts—$p_1\%$ and $p_2\%$ where $p_1 + p_2 < 1$, and tests whether these two parts of the chain have the same mean. Geweke et al. (1991) detail the calculation of the standard error for this test, a procedure known as tapering. After discarding the burn-in period, the traces of the parameters are divided into two even parts, and Table 8 reports p-values of the test with tapering of 4%, 8%, and 15%. A small p-value indicates the rejection of the null that the two parts of the chain have equal means.

The Geweke test results suggest that most parameters have converged, but there are some exceptions. It is also noticeable that although posterior means of some parameters are significantly greater than 0—suggesting self/cross-excitation behavior, their traces have not converged (e.g., $\tilde{\gamma}_{p-}^{\nu}$). We also perform a Raftery–Lewis test, and the results are similar.

Table 5 Log Bayes Factor (BF) of $\mathcal{M}_1$ to $\mathcal{M}_7$

AAPL						
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	3276.2	1768.5	3154.4	3205.5	2130.3	700.9
$\mathcal{M}_2$	2575.3	1067.6	2453.5	2504.6	1429.4	
$\mathcal{M}_3$	1145.9	−361.8	1024.1	1075.2		
$\mathcal{M}_4$	70.8	−1436.9	−51.1			
$\mathcal{M}_5$	121.9	−1385.9				
$\mathcal{M}_6$	1507.7					

KO						
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	4265.4	2351.4	3814.0	4212.8	2904.8	1017.7
$\mathcal{M}_2$	3247.7	1333.7	2796.3	3195.1	1887.0	
$\mathcal{M}_3$	1360.7	−553.3	909.3	1308.1		
$\mathcal{M}_4$	52.6	−1861.4	−398.8			
$\mathcal{M}_5$	451.4	−1462.6				
$\mathcal{M}_6$	1914.0					

BA						
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	10135.6	8409.4	9748.3	10167.9	9265.3	1559.2
$\mathcal{M}_2$	8576.4	6850.1	8189.1	8608.7	7706.1	
$\mathcal{M}_3$	870.3	−856.0	483.0	902.6		
$\mathcal{M}_4$	−32.3	−1758.5	−419.6			
$\mathcal{M}_5$	387.3	−1339.0				
$\mathcal{M}_6$	1726.2					

SPX						
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	7143.3	4159.2	6225.2	6487.5	4032.2	1010.0
$\mathcal{M}_2$	6133.3	3149.2	5215.2	5477.5	3022.1	
$\mathcal{M}_3$	3111.2	127.0	2193.0	2455.3		
$\mathcal{M}_4$	655.9	−2328.3	−262.3			
$\mathcal{M}_5$	918.1	−2066.0				
$\mathcal{M}_6$	2984.1					

JPM						
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	4484.9	1342.3	3814.9	4354.4	1056.1	337.7
$\mathcal{M}_2$	4147.2	1004.6	3477.2	4016.6	718.4	
$\mathcal{M}_3$	3428.8	286.2	2758.8	3298.2		
$\mathcal{M}_4$	130.5	−3012.1	−539.4			
$\mathcal{M}_5$	669.9	−2472.6				
$\mathcal{M}_6$	3142.6					

Note: The table presents the log BF of $\mathcal{M}_7$ to $\mathcal{M}_2$ (row) against $\mathcal{M}_1$ to $\mathcal{M}_6$ (column).

Table 6 Prior settings in the sensitivity analysis

	Original setting
Pr_1 : $\beta_\nu \sim U(0,1)$	$\beta_\nu \sim N(0,1)1_{\beta_\nu>0}$
Pr_2 : $\sigma_P^2 \sim IG(10,80)$	$\sigma_P^2 \sim IG(10,40)$
Pr_3 : $\boldsymbol{\vartheta} \sim \exp(0.1)$	$\boldsymbol{\vartheta} \sim N(0,0.1)1_{\boldsymbol{\vartheta}>0}$
Pr_4 : $\boldsymbol{\vartheta} \sim \exp(0.1), \boldsymbol{\beta} \sim \exp(0.1)$	$\boldsymbol{\vartheta} \sim N(0,0.1)1_{\boldsymbol{\vartheta}>0}, \boldsymbol{\beta} \sim N(0,0.1)1_{\boldsymbol{\beta}>0}$

The results are confined to Appendix C. We also provide the trace plots and posterior histograms of parameters in the Supplementary Appendix.

4 Clusters and Simulation Test

We next evaluate the models by focusing on their ability to capture features of the data of particular interest. We do this by simulating data from the estimated models and seeing whether these data have the features that characterize the actual data. This approach has been used by Hess and Iwata (1997) and Clements and Krolzig (2004) to assess whether a number of time-series models can reproduce business cycle features. Our approach is in line with that used in Clements and Krolzig (2004).

4.1 Cluster

In this section, we explain how we define a cluster of jumps. The way we define a cluster of jumps is similar to Foschi, Lilla, and Mancini (2019). The intuition is that jumps may escalate their underlying intensity to be above a "normal" level for a short period, before the intensity returns to the "normal" level. For example, suppose k jumps occurred, and these k jumps are a cluster. If $k = 1$, there are no jumps in the cluster other than the original one. In this case, the intensity returns to normal before any other jumps occur. Jumps occurring subsequent to this are assumed to belong to a new cluster. The formal definition is as follows.

Definition 1 (Cluster) *Given the occurrence of k jumps at times $t_1, \ldots, t_k$ during the period $[t_1, T]$ starting with an immigrant jump at t_1, these k jumps form a cluster if the intensity: (i) before the occurrence of t_1 jump, is within a range around ground intensity $\lambda_{t_1-} \in [\lambda_0, \tilde{\lambda}]$ and (ii) above the threshold $\tilde{\lambda}$ during the period $[t_1, T]$. The $\tilde{\lambda}$ is the tolerance level.*

4.2 Nonparametric Estimation

The models provide estimates of latent jumps and parameters from which the underlying intensity of jumps can be calculated, allowing us to determine the average number of jumps in a cluster. However, this would make the determination of the degree of clustering dependent on the model, and would favor those models that, by design, incorporate this feature over others which do not. To enable a fair comparison between the models, we estimate the feature of interest in the models' simulated output, not as a function of the models' parameters. We adopt a nonparametric estimation of price and variance jumps, and underlying jump intensities.

Table 7 Prior sensitivity analysis

	$\mathcal{M}_4$									
	Pr_0		Pr_1		Pr_2		Pr_3		Pr_4	
	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.
$\mu(*10^{-4})$	6.508	1.491	6.477	1.557	6.538	1.571	6.539	1.307	6.478	1.499
$\alpha_v(*10^{-3})$	0.167	0.006	0.169	0.006	0.166	0.005	0.170	0.005	0.169	0.005
β_v	−0.014	0.001	−0.013	−0.001	−0.013	0.001	−0.013	0.000	−0.014	0.001
$\sigma_V^2(*10^{-3})$	0.268	0.008	0.268	0.011	0.263	0.010	0.266	0.004	0.265	0.012
ρ	−0.240	0.011	−0.236	0.011	−0.240	0.010	−0.235	0.011	−0.244	0.012
μ_P	−0.032	0.019	−0.036	0.020	−0.037	0.024	−0.030	0.016	−0.035	0.021
σ_P	0.160	0.018	0.161	0.019	0.161	0.017	0.158	0.018	0.161	0.018
μ_V	0.005	0.002	0.005	0.002	0.005	0.002	0.005	0.002	0.005	0.002
$\boldsymbol{\vartheta}$	0.450	0.324	4.8%	13.1%	7.3%	15.8%	5.1%	20.0%	5.2%	15.7%
$\tilde{\boldsymbol{\beta}}$	0.351	0.279	3.1%	16.8%	5.9%	18.3%	4.8%	16.2%	5.4%	19.8%
$\tilde{\gamma}$	0.053	0.031	3.5%	13.7%	2.3%	8.8%	5.0%	15.1%	4.9%	14.6%
log-likelihood	253,591		253,548		253,576		253,491		253,453	
DIC	491,936		492,301		491,953		491,849		492,391	

(continued)

Table 7 Continued

	$\mathcal{M}_7$									
	Pr_0		Pr_1		Pr_2		Pr_3		Pr_4	
	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.	Posterior mean	Posterior std. dev.
$\mu(*10^{-4})$	6.610	1.815	6.657	1.920	6.572	1.941	6.640	1.784	6.602	1.732
$\alpha_v(*10^{-3})$	0.171	0.005	0.175	0.006	0.171	0.006	0.171	0.005	0.168	0.005
β_v	−0.014	0.001	−0.014	0.002	−0.013	0.001	−0.014	−0.001	−0.013	0.001
$\sigma_V^2(*10^{-3})$	0.272	0.008	0.274	0.015	0.274	0.009	0.275	0.011	0.274	0.013
ρ	−0.237	0.010	−0.232	0.009	−0.239	0.010	−0.236	0.010	−0.241	0.009
μ_P	−0.039	0.015	−0.040	0.013	−0.042	0.016	−0.038	0.021	−0.041	0.014
σ_P	0.159	0.017	0.157	0.019	0.157	0.018	0.159	0.017	0.160	0.017
μ_V	0.005	0.002	0.005	0.002	0.005	0.002	0.005	0.002	0.005	0.002
$\boldsymbol{\vartheta}$	0.507	0.157	6.8%	18.0%	4.1%	10.2%	3.8%	10.8%	3.0%	19.5%
$\tilde{\boldsymbol{\beta}}$	0.811	0.240	4.4%	14.4%	4.2%	11.9%	4.0%	15.0%	3.7%	14.5%
log-likelihood	254,246		254,002		254,272		254,216		254,016	
DIC	490,330		490,562		490,397		490,406		490,576	

Notes: Since $\{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}\}$ are 3×3 dimensions (i.e., $\tilde{\boldsymbol{\beta}} = \tilde{\beta}_q^i$, $i, q \in \{p+, p-, v\}$), in this table, we report the mean percentage of deviations from the Pr_0 and full details are provided in the Supplementary Appendix.

Table 8 Geweke diagnostic test

	4%	8%	15%		4%	8%	15%
$\mu(*10^{-4})$	0.011	0.032	0.049	MMHP parameters in $\mathcal{M}_4$			
$\alpha(*10^{-3})$	0.205	0.326	0.393	$\tilde{\beta}_{p+}^{p+}$	0.499	0.593	0.641
β	0.218	0.331	0.390	$\tilde{\beta}_{p-}^{p+}$	0.357	0.421	0.533
$\sigma_V^2(*10^{-3})$	0.482	0.560	0.593	$\tilde{\beta}_{v}^{p+}$	0.002	0.023	0.074
ρ	0.059	0.100	0.120	$\tilde{\beta}_{p+}^{p-}$	0.003	0.023	0.048
μ_P	0.562	0.627	0.673	$\tilde{\beta}_{p-}^{p-}$	0.099	0.181	0.323
σ_P	0.508	0.581	0.595	$\tilde{\beta}_{v}^{p-}$	0.027	0.102	0.196
μ_V	0.349	0.343	0.263	$\tilde{\beta}_{v}^{v}$	0.000	0.004	0.017
				$\tilde{\beta}_{p+}^{v}$	0.049	0.094	0.109
MMHP parameters in $\mathcal{M}_7$				$\tilde{\beta}_{p-}^{v}$	0.489	0.597	0.662
$\tilde{\beta}_{p+}^{p+}$	0.432	0.508	0.597	$\tilde{\gamma}_{p+}^{p+}$	0.687	0.644	0.600
$\tilde{\beta}_{p-}^{p+}$	0.596	0.681	0.730	$\tilde{\gamma}_{p-}^{p+}$	0.121	0.136	0.112
$\tilde{\beta}_{v}^{p+}$	0.601	0.650	0.638	$\tilde{\gamma}_{v}^{p+}$	0.000	0.000	0.000
$\tilde{\beta}_{p+}^{p-}$	0.445	0.554	0.613	$\tilde{\gamma}_{p+}^{p-}$	0.186	0.107	0.032
$\tilde{\beta}_{p-}^{p-}$	0.242	0.368	0.461	$\tilde{\gamma}_{p-}^{p-}$	0.792	0.804	0.811
$\tilde{\beta}_{v}^{p-}$	0.884	0.910	0.924	$\tilde{\gamma}_{v}^{p-}$	0.000	0.000	0.000
$\tilde{\beta}_{p+}^{v}$	0.027	0.076	0.107	$\tilde{\gamma}_{p+}^{v}$	0.173	0.185	0.205
$\tilde{\beta}_{p-}^{v}$	0.797	0.836	0.856	$\tilde{\gamma}_{p-}^{v}$	0.028	0.037	0.037
$\tilde{\beta}_{v}^{v}$	0.024	0.078	0.127	$\tilde{\gamma}_{v}^{v}$	0.047	0.112	0.132

Notes: The burn-in period is discarded and the remaining traces of parameters are divided into two even parts for the test. The table reports p-values of the Geweke test tapered with 4%, 8%, and 15% autocovariance. A small p-value indicates the rejection of the null that two parts of the chain have different means.

Assume the price process over the time interval $[0, T]$ is observed at times $t_i := i\Delta_n, 1 \le i \le n$, where $\Delta_n = T/n$ presents the time increments. We further denote $\Delta_i^n J := J_{t_i} - J_{t_{i-1}}$ as jumps.[5] We filtered price and variance jumps $(\Delta_i^n J^P, \Delta_i^n J^V)$ nonparametrically, and details can be seen in Appendix D. We use kernel density estimation (KDE) to estimate the underlying intensity of jumps, and then determine the clusters. The underlying intensity at a given time $\tau \in (0, T)$ is defined as:

$$\hat{\lambda}_\tau^q = \sum_{i=1}^{n} K_h(t_{i-1} - \tau)\Delta_i^n J^q, \qquad q \in \{P^+, P^-, V\} \tag{16}$$

where $K_h(\cdot)$ is a kernel function with bandwidth h, such that $\int K_h(x)dx = 1$ and $K_h(x) = \frac{1}{h}K(\frac{x}{h})$. We use the exponential kernel $K(x) = \frac{1}{2}\exp(-x)$ and set $h = 40\Delta_n$. In terms of the tolerance level, we take $\tilde{\lambda} = \frac{\zeta}{2h}e^{-\frac{1}{h}}$, $\zeta = 5\%$, which is calculated as 5% of the underlying intensity at the time that the immigrant jump occurred.

Setting $h = 40\Delta_n$ means that intensity is determined based on the trading day around the period, in keeping with our aim of capturing the intraday jump clustering effect.

5 $\Delta_i^n J$ is the same as ΔJ_t in Equations (7) and (8). But it is a more general notation for the convenience of demonstrating our nonparametric estimation approaches.

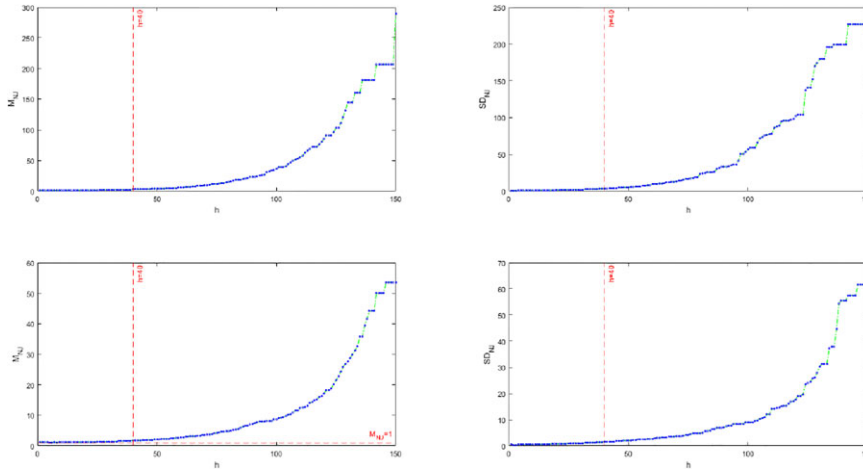

Figure 3 Mean and standard deviation of the number of jumps in clusters under different bandwidth.

Notes: We randomly generate 1000 events during $100{,}000\Delta_n$ ($\lambda_t = 0.01$) and count the number of jumps in clusters using the method we proposed. The upper two graphs plot the average number of jumps and the standard deviation of the number of jumps in clusters given different bandwidths. The bottom two figures plot the mean and standard deviation of the number of negative jumps in the S&P 500 Index with different bandwidths.

We also experiment with different bandwidths. We find that the choice of bandwidth does not unduly affect our results for a reasonable range of values.

We have investigated this by randomly generating 1000 events during $100{,}000\Delta_n$ ($\lambda_t = 0.01$) and counting the number of jumps in clusters using the proposed method. We plot the average number of jumps and the standard deviation of the number of jumps in clusters given different bandwidths in the upper two graphs of Figure 3. Ideally, the average number of jumps should equal 1 with a very low standard deviation since events are randomly drawn. The figure shows that results are very similar for bandwidths smaller than seventy, but become unreasonable for wider bandwidths. The bottom two figures plot the means and standard deviations of the number of negative jumps in S&P 500 Index with different bandwidths. Again, results are very similar, with bandwidths ranging from thirty to sixty. We conclude that our findings are not unduly sensitive to the choice of bandwidth.

Similarly, the choice of ζ in calculating the tolerance level ($\tilde{\lambda}$) does not unduly affect our results either. We plot the means and standard deviations of the number of negative return jump in clusters under different ζ in Figure 4. Very small values of ζ lead to a too-small tolerance level $\tilde{\lambda}$, and an unreasonably high average number of jumps in clusters. The reverse happens for large values of ζ. For a range of values between these extremes, we find the "right" average size of one.

4.3 Characteristics of Interest

We choose the average number of jumps in a cluster, and the standard deviation of the number of jumps in a cluster, as the two characteristics of interest. Clustering is a feature of

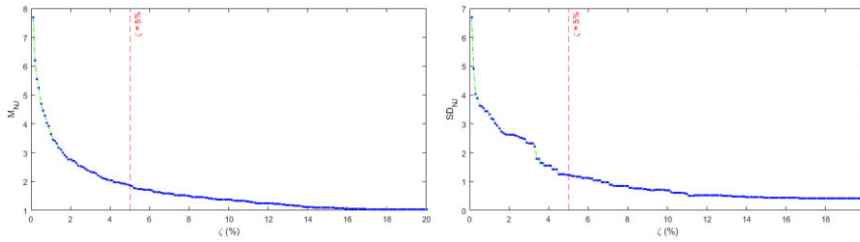

Figure 4 Mean and standard deviation of the number of jumps in clusters under different ζ.

Notes: This figure studies those statistics of negative return jumps in S&P 500 Index. We test a range of ζ from 0.1% to 20%.

the data, so we compare the models in this dimension—their ability to match this feature of the data—irrespective of their overall fit. Poisson processes treat occurrences of jumps as independent events, so there should, on average, be only one jump in each cluster, and we would expect such processes to fail to match the data in this regard. On the other hand, multivariate HP and MMHP emphasize the clustering of jumps, and are expected to produce clusters with more than one jump. However, unlike HP, MMHP allows jumps with larger sizes to raise intensities, which also take a longer time to die away. Hence the distribution of the number of jumps under HP and MMHP can be different, and may allow some discrimination between these two models.

We consider $\mathcal{M}_1$, $\mathcal{M}_3$, $\mathcal{M}_4$, and $\mathcal{M}_7$, and treat these models as the DGP, setting their parameters to their posterior means of the Bayesian model estimation. The feature value from the actual data can then be compared to the empirical distributions of the features from the models. If the feature value on the actual data is "extreme" relative to a model's distribution, we infer that the model is unable to capture that feature.

Specifically, denote by C_i the i^{th} characteristic (either the 1st or 2nd moments of the number of jumps). Assumes there are J simulations. We rank the characteristic on each replication: $\{C_i^{(1)} \dots C_i^{(J)}\}$. A confidence interval with a significance level SL is given by $[F_{C_i}^{-1}\left(\frac{SL}{2}\right), F_{C_i}^{-1}(1-\frac{SL}{2})]$, where F_{C_i} denotes the empirical cumulative distribution of characteristic C_i. For example, taking J=10,000, and significance level at 5%, the upper and lower limit of the confidence interval will then be $F_{C_i}^{-1}(2.5\%)$, and $F_{C_i}^{-1}(97.5\%)$, confidence interval, thus, is taken as $[C_i^{(250)}, C_i^{(9750)}]$.

4.4 Testing Results

Table 9 presents the summary statistics for the number of jumps in clusters. On average, there are two-to-three jumps per cluster, with more jumps in clusters for some individual stock data. In addition, there is a clear pattern that the average number of variance jumps is lower than that of price jumps. Also, around half of the clusters consist of more than one jump. We also collect the time spans that clusters of jumps cover, and denote these as cluster lengths. As can be seen in the table, the cluster length of negative return jumps ranges from thirty to seventy time intervals, that is, 2.5–6 hours. A negative return jump can impact jump probabilities for the next 2.5–6 hours. It is also apparent that negative return jump clusters last longer than those positive return jumps and variance jumps.

Table 9 Summary statistics of clusters

		No. of cluster	M_{NJ}	SD_{NJ}	Max(NJ)	M_L	Max(L)	No. of cluster containing only one jump
AAPL	$p+$	660	3.936	3.406	24	41	303	166
	$p-$	663	4.394	3.988	36	68	402	167
	v	198	1.480	0.900	6	12	53	138
BA	$p+$	670	3.031	2.547	16	42	318	229
	$p-$	645	3.093	2.751	22	68	425	222
	v	213	1.268	0.574	4	11	55	168
JPM	$p+$	517	1.994	1.443	10	29	284	272
	$p-$	518	2.073	1.572	12	41	192	261
	v	117	1.265	0.621	4	10	31	95
KO	$p+$	432	1.672	0.921	7	18	117	265
	$p-$	391	1.714	1.135	9	36	275	231
	v	99	1.192	0.444	3	18	47	82
SPX	$p+$	313	1.454	0.905	7	28	119	221
	$p-$	378	1.870	1.230	9	36	123	198
	v	198	1.260	0.587	4	9	48	179

Notes: M_{NJ} denotes mean number of jumps in clusters, SD_{NJ} denotes standard deviation of number of jumps in clusters, max(NJ) denotes maximum number of jumps in a cluster. M_L denotes the mean of cluster length (length of a cluster of jumps cover) and $Max(L)$ denotes the maximum value of cluster length. Cluster length figures are rounded to integers.

We simulate 50,000 series of data ($J = 50{,}000$) and each contains 100,000 data points, Table 10 reports the results from each of the models ($\mathcal{M}_1$, $\mathcal{M}_3$, $\mathcal{M}_4$, $\mathcal{M}_7$). In terms of price jumps, the HP model ($\mathcal{M}_3$) is clearly unable to reproduce the features, while the MMHP-class models ($\mathcal{M}_4$ and $\mathcal{M}_7$) perform better. However, in S&P 500 (SPX) data, $\mathcal{M}_7$ outperforms $\mathcal{M}_4$.

In the case of variance jumps, the results are different. For example, for the BA and KO data, the HP model ($\mathcal{M}_3$) is able to adequately capture the clustering feature. This is likely due to the smaller average number of jumps, and standard deviation, for variances compared to prices. These results are in line with the empirical results in Section 3 (see Table 3), where variance jumps are less likely impacted by their mark values or jump sizes. It is noticeable that, although $\mathcal{M}_4$ and $\mathcal{M}_7$ are both MMHP-type models, $\mathcal{M}_4$ always provides higher estimates of the two features. Note we also conduct the test with ETF data, and the results are consistent. We provide details of these results in the Supplementary Appendix.

5 Conclusion

In this article, we propose a dynamic bivariate jump-diffusion process, in which jump intensities are modeled by a three-dimensional MHP to allow the occurrences and sizes of jumps to affect future intensities and, thus, capture the clustering features. Unlike other stochastic volatility state-space models that apply daily data, we use intraday high-frequency data.

Table 10 Simulation results

Characteristics	Positive price jump					Negative price jump					Variance jump				
AAPL	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	3.936	1.072**	1.213**	4.371	3.711	4.394	1.059**	1.341**	5.924†	3.957	1.48	1.035**	1.069**	1.862*	1.321
SD_{NJ}	3.406	0.211**	0.318**	3.128	2.962	3.988	0.392**	0.429**	3.749	3.152	0.9	0.203**	0.266**	1.159†	0.671
BA	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	3.031	1.114**	1.428**	4.447*	3.641	3.093	1.118**	1.466**	3.367	2.618	1.268	1.032**	1.172	1.461**	1.231
SD_{NJ}	2.547	0.365**	0.495**	2.616	2.353	2.751	0.371**	0.395**	2.626	2.317	0.574	0.173**	0.333*	1.072**	0.645
JPM	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	1.994	1.07**	1.367**	3.105*	2.286	2.073	1.038**	1.313**	2.342	2.059	1.265	1.032**	1.066**	1.372	1.218
SD_{NJ}	1.443	0.273**	0.482**	2.047†	1.549	1.572	0.194**	0.641**	1.804	1.366	0.621	0.165**	0.263**	0.474†	0.692
KO	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	1.672	1.061**	1.344**	1.901	1.693	1.714	1.072**	1.328**	1.917	1.896	1.192	1.047*	1.159	1.385*	1.268†
SD_{NJ}	0.921	0.295**	0.211**	0.863	0.933	1.135	0.345**	0.469**	1.363	1.117	0.444	0.334	0.346	0.791*	0.268*
SPX	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	1.454	1.062**	1.129**	1.351	1.416	1.87	1.094**	1.284**	2.129	1.85	1.258	1.029**	1.164*	1.48**	1.254
SD_{NJ}	0.905	0.249**	0.372**	0.683†	0.923	1.23	0.396**	0.502**	1.408	1.371	0.587	0.28**	0.356*	1.046**	0.67

Notes: M_{NJ} denotes the mean of the number of jumps in clusters, and SD_{NJ} denotes the standard deviation of the number of jumps in clusters.

†Indicates less than 10 p.c. of simulations were further out in the tails than the sample estimate.

*Indicates less than 5 p.c. of simulations were further out in the tails than the sample estimate.

**Indicates less than 1 p.c. of simulations were further out in the tails than the sample estimate.

Table 11 Priors specification

General parameters		$\mathcal{M}_4$'s impact function	
μ	$N(0,25)$	$\tilde{\alpha}$	$N(0,0.2)\mathbf{1}_{\tilde{\alpha}>0}$
α_v	$N(0,1)\mathbf{1}_{\alpha_v>0}$	$\tilde{\beta}$	$N(0,1)\mathbf{1}_{\tilde{\beta}>0}$
β_v	$N(0,1)\mathbf{1}_{\beta_v>0}$	$\tilde{\gamma}$	$N(0,1)\mathbf{1}_{\tilde{\gamma}>0}$
f_τ	$N(0,0.5)\mathbf{1}_{f_\tau>0}$		
σ_V	$IG(2.5,0.1)$	$\mathcal{M}_5$'s impact function	
ρ	$U(-\mathbf{1},\mathbf{1})$	$\tilde{\alpha}$	$N(0,0.3)\mathbf{1}_{\tilde{\alpha}>0}$
ξ^P	$N(0,50)$		
ξ^V	$N(0,10)\mathbf{1}_{\xi^V>0}$	$\mathcal{M}_6$'s impact function	
μ_P	$N(0,50)$	$\tilde{\alpha}$	$N(0,30)\mathbf{1}_{\tilde{\alpha}>0}$
σ_P^2	$IG(10,40)$		
μ_V	$IG(10,20)$	$\mathcal{M}_7$'s impact function	
$\boldsymbol{\vartheta}$	$N(0,0.1)\mathbf{1}_{\boldsymbol{\vartheta}>0}$	$\tilde{\alpha}$	$N(0,0.2)\mathbf{1}_{\tilde{\alpha}>0}$
β	$N(0,0.1)\mathbf{1}_{\beta>0}$	$\tilde{\beta}$	$N(0,1)\mathbf{1}_{\tilde{\beta}>0}$

Note: The table presents priors settings in the MCMC algorithm of parameter estimations.

Table 12 Raftery–Lewis Test

	N_1	N_2	$I-stat$		N_1	N_2	$I-stat$
$\mu(*10^{-4})$	6	1250	2.1	MMHP parameters in $\mathcal{M}_4$			
$\alpha_v(*10^{-3})$	87	14,210	23.8	$\tilde{\beta}_{p+}^{p+}$	40	6490	10.9
β_v	26	4674	7.8	$\tilde{\beta}_{p-}^{p+}$	50	8244	13.8
$\sigma_V^2(*10^{-3})$	312	56,567	94.8	$\tilde{\beta}_{v}^{p+}$	59	10,269	17.2
ρ	76	12,813	21.5	$\tilde{\beta}_{p+}^{p-}$	168	26,356	44.2
μ_P	188	31,863	53.4	$\tilde{\beta}_{p-}^{p-}$	308	44,283	74.3
σ_P	228	39,136	65.6	$\tilde{\beta}_{v}^{p-}$	106	14,858	24.9
μ_V	11	2235	3.7	$\tilde{\beta}_{p+}^{v}$	48	7785	13.1
				$\tilde{\beta}_{p-}^{v}$	80	16,368	27.4
MMHP parameters in $\mathcal{M}_7$				$\tilde{\beta}_{v}^{v}$	114	18,612	31.2
$\tilde{\beta}_{p+}^{p+}$	186	29,298	49.1	$\tilde{\gamma}_{p+}^{p+}$	14	2748	4.6
$\tilde{\beta}_{p-}^{p+}$	168	27,405	46.0	$\tilde{\gamma}_{p-}^{p+}$	12	2208	3.7
$\tilde{\beta}_{v}^{p+}$	51	8400	14.1	$\tilde{\gamma}_{v}^{p+}$	190	28,938	48.5
$\tilde{\beta}_{p+}^{p-}$	91	16,548	27.7	$\tilde{\gamma}_{p+}^{p-}$	12	2172	3.6
$\tilde{\beta}_{p-}^{p-}$	213	33,992	57.0	$\tilde{\gamma}_{p-}^{p-}$	12	2274	3.8
$\tilde{\beta}_{v}^{p-}$	69	11,760	19.7	$\tilde{\gamma}_{v}^{p-}$	256	39,952	67.0
$\tilde{\beta}_{p+}^{v}$	73	13,533	22.7	$\tilde{\gamma}_{p+}^{v}$	13	2334	3.9
$\tilde{\beta}_{p-}^{v}$	77	11,892	19.9	$\tilde{\gamma}_{p-}^{v}$	11	2076	3.5
$\tilde{\beta}_{v}^{v}$	72	13,512	22.6	$\tilde{\gamma}_{v}^{v}$	291	47,003	78.8

Notes: Test settings are quantile $q = 0.025$, accuracy $r = 0.0125$, probability $s = 0.95$. N_1: number of iterations that should be discarded in the burn-in period. N_2: least number of iterations for parameters to converge. $I-stat$: the extent of autocorrelation of the parameter's chain.

In addition, we employ an intraday periodic component in the process. A Bayesian MCMC algorithm is constructed to jointly estimate parameters and latent variables in the model.

We find evidence of strong intraday jump clustering in our empirical study. We find that self-excitation tends to be hidden by periodicity. We quantify the changing patterns of jump intensity, and show that jump intensity can rise to over 0.15 after the occurrence of large jumps. We investigate the interactions between positive price jumps, negative price jumps, and variance jumps. These are apparent in the branching coefficient matrix $\boldsymbol{\vartheta}$, which represents the extent to which jumps in one dimension affect intensity in other dimensions. This turns out to imply some asymmetries. Further, our results from the MHP suggest that the extent to which jumps inflate future intensities is positively correlated with jump sizes: large jumps tend to escalate the probability of jumps in the near future. We assess the fit of the various models via DIC and the Bayes factor, and find that modeling jumps by the MHP are preferred by these criteria.

In addition, we study the numbers of jumps in clusters, in price, and variance, and we find a cluster of jumps can cover 2.5–6 hours on average. Using the mean and variance of the number of jumps in clusters, we consider the ability of the models to reproduce these two characteristics, by simulating artificial data from the models. The MMHP models generally outperform other models, although the simpler HP model is able to capture the properties of variance jumps.

Our study indicates the various novel features of our models—including allowing the occurrences and sizes of jumps returns and variances to affect future intensities—can be valuable in fitting the data and matching certain key characteristics of the data. A possible area for future research would be to extend this approach to analyze risk premiums using options data.

Supplemental Data

Supplemental data is available at https://www.datahostingsite.com.

Appendix A. Bayesian MCMC Algorithm and Specification of Priors

In order to obtain the posterior distribution in Equation (14), we randomly sample from a set of conditional posteriors derived from Bayes's rule. An MCMC algorithm is constructed to approximate the posterior implied by the estimated model. Therefore, for $i = 1, 2, \ldots, n, \ldots, N$:

1. Sample static parameters

   Draw $\Theta_1^{(i)}$ from $p\left(\Theta_1^{(i)} | P_t, \Theta_2^{(i-1)}, \Theta_3^{(i-1)}, \ldots, \Theta_k^{(i-1)}, \Omega_t^{(i-1)}\right)$,

   $\vdots$

   Draw $\Theta_k^{(i)}$ from $p\left(\Theta_k^{(i)} | P_t, \Theta_1^{(i-1)}, \Theta_2^{(i-1)}, \ldots, \Theta_{k-1}^{(i-1)}, \Omega_t^{(i-1)}\right)$.

2. Sample jumps and jump sizes
   for $t = 1, 2, \ldots, T$:

   Draw $\Delta J_t^{P(i)}$ from $p\left(\Delta J_t^{P(i)} | P_t, \Theta^{(i)}, \xi_t^{P(i-1)}, \Delta J_t^{V(i-1)}, \xi_t^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}, V^{(i-1)}\right)$,

   Draw $\xi_t^{P(i)}$ from $p\left(\xi_t^{P(i)} | P_t, \Theta^{(i)}, \Delta J_t^{P(i)}, \Delta J_t^{V(i-1)}, \xi_t^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}, V_t^{(i-1)}\right)$,

   Draw $\Delta J_t^{V(i)}$ from $p\left(\Delta J_t^{V(i)} | P_t, \Theta^{(i)}, \xi_t^{P(i)}, \Delta J_t^{P(i)}, \xi_t^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}, V^{(i-1)}\right)$,

   Draw $\xi_t^{V(i)}$ from $p\left(\xi_t^{V(i)} | P_t, \Theta^{(i)}, \xi_t^{P(i)}, \Delta J_t^{P(i)}, \Delta J_t^{V(i)}, \boldsymbol{\vartheta}^{(i-1)}, V^{(i-1)}\right)$.

3. Sample variance
   for $t = 1, 2, \ldots, T$:

   Draw V_t from $p\left(V_t^{(i)} | P_t, \Theta^{(i)}, \xi_t^{P(i)}, \Delta J_t^{P(i)}, \xi_t^{V(i)}, \Delta J_t^{V(i)}, \boldsymbol{\vartheta}^{(i-1)}\right)$,

4. Sample Branching Coefficient Matrix

   for $t = 1, 2, \ldots, T$:

   Draw $\boldsymbol{\vartheta}^{(i)}$ from $p\left(\boldsymbol{\vartheta}^{(i)} | P_t, \Theta^{(i)}, \xi_t^{P(i)}, \Delta J_t^{P(i)}, \xi_t^{V(i)}, \Delta J_t^{V(i)}, V_t^{(i)}\right)$

We set $N = 40,000$ as the total number of iterations and $n = 10,000$ as the burn-in period, which will be discarded. For those conditional posterior distributions where corresponding conjugate priors can be found and posteriors can be obtained in closed form, we adopt Gibbs sampling; for those posteriors that are unknown, we use Metropolis–Hastings (MH) to approximate posteriors. MH involves drawing a sample from a proposal density and another random number from a uniform distribution to decide whether the proposal draw should be accepted or rejected. Ultimately, we specify priors in our model as follows (Table 11).

Moreover, when we run the MCMC algorithm, the original data and estimated data are amplified by 10,000 times.

Appendix B. Marginal Likelihood of Models

We compute the marginal likelihood of data given different models as follows:

$$p(P_t|\mathcal{M}_i) = \frac{p(P_t|\Theta_i, \mathcal{M}_i)p(\Theta_i|\mathcal{M}_i)}{p(\Theta_i|P_t, \mathcal{M}_i)}, \tag{IB.17}$$

where $i = 1, \ldots, 7$ denotes seven different models, $\mathcal{M}_i$ denotes corresponding static parameters. The likelihood of data given models and static parameters is further marginalized over k latent variables $\Omega_t^{(k)}$ as follows:

$$p(P_t|\Theta_i,\mathcal{M}_i)=\int p\left(P_t|\Omega_t^{(k)},\Theta_i,\mathcal{M}_i\right)p\left(\Omega_t^{(k)}|\Theta_i,\mathcal{M}_i\right)d\Omega_t^{(k)}, \tag{IB.18}$$

Using the output of previous MCMC outputs, this is the marginal likelihood of the data and models averaged over the latent variables. Similarly, the conditional posterior of static parameters is also marginalized over latent variables:

$$p(\Theta_i|P_t,\mathcal{M}_i)=\int p\left(\Theta_i|P_t,\Omega_t^{(k)},\mathcal{M}_i\right)d\Omega_t^{(k)}, \tag{IB.19}$$

Following methods proposed by Chib (1995) and Chib and Jeliazkov (2001), we decompose the static parameter vector Θ_i into two components: $\theta_{1,i}$ denoting parameters in the price and variance processes, and $\theta_{2,i}$, denoting parameters in jumps components (Hawkes kernel). Therefore,

$$p(\Theta_i|P_t,\mathcal{M}_i)=p(\theta_{1,i}|P_t,\mathcal{M}_i)p(\theta_{2,i}|\theta_{1,i},P_t,\mathcal{M}_i), \tag{IB.20}$$

The likelihood of a multivariate HP is derived in Liniger (2009).

Appendix C. Raftery–Lewis Test

Raftery–Lewis test, proposed by Raftery and Lewis (1991), provides bounds of the number of iterations that should be run and the number that should be discarded given a quantile that we wish to estimate with a desired accuracy and associated probability.

In this article, we specify the quantile $q = 0.025$, accuracy $r = 0.0125$, and probability $s = 0.95$. The test results are provided in Table 12. N_1 denotes the number of iterations that should be discarded in the burn-in period. N_2 denotes the least number of iterations for parameters to converge. $I - stat$ presents the extent of autocorrelation of the parameter's chain. Note we set our MCMC number of iterations as 40,000 with the first 10,000 as the burn-in period. Overall, most parameters have converged as indicated by the results. Some parameters appear to have moderate autocorrelation.

Appendix D. Nonparametric Estimation

In this section, we discuss our estimation approaches for the variance and jumps $(V_i^n, \Delta_i^n J^P, \Delta_i^n J^V)$.[6] These estimations are used in deciding jump intensities.

D.1 High-frequency Variance Estimator and Price Jump Detection

Assume the price process over the time interval $[0, T]$ is observed at times $t_i := i\Delta_n, 1 \le i \le n$, where $\Delta_n = T/n$ presents the time increments. Thus, we have $\Delta_i^n P := P_{t_i} - P_{t_{i-1}}$ as asset returns, and $\Delta_i^n P' = \Delta_i^n P / S_{t_i}$ as asset returns without intraday periodicity. The estimation of the intraday periodicity component S_{t_i} will be introduced in Section D.3.

6 For the convenience of demonstrating our estimations, we use these notations, but they are the same as $V_t, \Delta J_t^P, \Delta J_t^V$, respectively, in Equations (7) and (8).

Our nonparametric jump filtering method is mainly based on Mancini, Mattiussi, and Renò (2015) and Figueroa-López and Mancini (2019). We identify a jump at time t, $J_t = 1$, when the squared return is greater than a threshold, $(\Delta_i^n P')^2 > \hat{V}_{t_i}^2 \cdot 2\Delta_n \log \frac{1}{\Delta_n}$. $\hat{V}_{t_i}^2$ is a nonparametric estimator of spot variance based on pretruncated returns:

$$\hat{V}_\tau^2 = \sum_{i=1}^{n} f_h(t_i - \tau)(\Delta_i^n P')^2 \cdot \mathbf{1}_{\{(\Delta_i^n P')^2 \leq 9\Delta_n^{0.99}\}} \tag{ID.21}$$

where $f_h(\cdot)$ is weight function, $f_h(t) = \frac{1}{h} \cdot \frac{e^{-|t/h|}}{2}$. The idea of this filtering is to extract those standardized squared returns $((\Delta_i^n P')^2 / \hat{V}_{t_i}^2)$ which are not likely to have been generated by a Brownian motion, given the assumed threshold, that is, that their absolute value is greater than the threshold of $\sqrt{2\log(1/\Delta_n)}$.

D.2 Variance Jumps Detection

In testing variance jumps, we set our null hypothesis $H_0 : |\hat{V}_{t_i}^2 - \hat{V}_{t_i-}^2| = 0$ against the alternative hypothesis that there is a variance jump $|\hat{V}_{t_i}^2 - \hat{V}_{t_i-}^2| > 0$. So, the test statistic in this test should be a function of the difference of spot variance $f(\hat{V}_{t_i}^2, \hat{V}_{t_i-}^2)$. Following Jacod et al. (2010), we construct our test statistics as follows:

$$\mathcal{L}^v(t_i) = 2\log\left(\frac{1}{2}\left(\hat{V}_{t_i}^2 + \hat{V}_{t_i-}^2\right)\right) - log\left(\hat{V}_{t_i}^2\right) - log\left(\hat{V}_{t_i-}^2\right) \tag{ID.22}$$

and $n^b \mathcal{L}^v(t_i) \to \mathcal{X}_1^2$, where n denotes the number of observations and $b = \frac{1}{2} - \Delta_n$. The temporal variance estimator $\hat{V}_{t_i}^2$ is specified in Equation (ID.21).

D.3 Intraday Periodic Effects

Boudt, Croux, and Laurent (2011) show that taking intraday periodicity into account can improve the overall accuracy of jump detection. We adopt a weighted standard deviation (WSD) estimator proposed by Boudt, Croux, and Laurent (2011), which is based on a shortest half-scale estimator proposed by Rousseeuw and Leroy (1988). Using this approach, we define the order statistics of returns $\bar{r}_{(1),i} \leq \bar{r}_{(2),i} \leq \ldots \leq \bar{r}_{(T_i),i}$. The shortest half-scale statistics are determined as follows:

$$ShortH_i = 0.741 \cdot min\{\bar{r}_{(h_i),i} - \bar{r}_{(1),i}, \ldots, \bar{r}_{(T_i),i} - \bar{r}_{(T_i - h_i + 1),i}\}, \tag{ID.23}$$

where $h_i = \frac{T_i}{2} + 1$, which makes the statistics essentially minimum differences among all of the return's halves. The shortest half-scale estimator for periodicity is given by:

$$\hat{f}_i^{ShortH} = \frac{ShortH_i}{\frac{1}{T}\sum_{j=1}^{T} ShortH_j^2}, \tag{ID.24}$$

where T denotes the number of observations within a day and the WSD estimator can be obtained as follow:

$$\hat{f}_i^{WSD} = \frac{WSD_i}{\frac{1}{T}\sum_{j=1}^{T} WSD_j^2}, \tag{ID.25}$$

where:

$$WSD_j = \sqrt{1.081 \cdot \frac{\sum_{k=1}^{T_j} w_{k,j}\bar{r}_{k,j}^2}{\sum_{k=1}^{T_j} w_{k,j}}}, \tag{ID.26}$$

$w_{k,j} = I(\bar{r}_{k,j}/\hat{f}_i^{ShortH})$ is a weight function with a identification function $I(\cdot)$, such that $I(x) = 1$ if $x \leq 6.635$ and 0 otherwise. We use the WSD estimator ($\hat{f}_i^{WSD}$) as a periodic component (S_{t_i}) estimator.

References

Aït-Sahalia, Y., J. Cacho-Diaz, and R. J. Laeven. 2015. Modeling Financial Contagion Using Mutually Exciting Jump Processes. *Journal of Financial Economics* 117: 585–606.

Ait-Sahalia, Y., J. Fan, and Y. Li. 2013. The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency. *Journal of Financial Economics* 109: 224–249.

Andersen, T. G., and T. Bollerslev. 1997. Intraday Periodicity and Volatility Persistence in Financial Markets. *Journal of Empirical Finance* 4: 115–158.

Asgharian, H., and C. Bengtsson. 2006. Jump Spillover in International Equity Markets. *Journal of Financial Econometrics* 4: 167–203.

Bandi, F. M., and R. Renò. 2016. Price and Volatility Co-Jumps. *Journal of Financial Economics* 119: 107–146.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2008. Designing Realized Kernels to Measure the Ex Post Variation of Equity Prices in the Presence of Noise. *Econometrica* 76: 1481–1536.

Barndorff-Nielsen, O. E., and N. Shephard. 2006. Impact of Jumps on Returns and Realised Variances: Econometric Analysis of Time-Deformed Lévy Processes. *Journal of Econometrics* 131: 217–252.

Bollerslev, T., S. Z. Li, and B. Zhao. 2020. Good Volatility, Bad Volatility, and the Cross Section of Stock Returns. *Journal of Financial and Quantitative Analysis* 55: 751–781.

Boudt, K., C. Croux, and S. Laurent. 2011. Robust Estimation of Intraweek Periodicity in Volatility and Jump Detection. *Journal of Empirical Finance* 18: 353–367.

Chib, S. 1995. Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* 90: 1313–1321.

Chib, S., and I. Jeliazkov. 2001. Marginal Likelihood from the Metropolis–Hastings Output. *Journal of the American Statistical Association* 96: 270–281.

Chib, S., F. Nardari, and N. Shephard. 2002. Markov Chain Monte Carlo Methods for Stochastic Volatility Models. *Journal of Econometrics* 108: 281–316.

Clements, M. P., and H.-M. Krolzig. 2004. Can Regime-Switching Models Reproduce the Business Cycle Features of Us Aggregate Consumption, Investment and Output? *International Journal of Finance & Economics* 9: 1–14.

Daley, D. J., and D. Vere-Jones. 2003. *An introduction to the theory of point processes, volume 1: Elementary theory and methods*. New York Berlin Heidelberg: Springer 11, 69.

Duffie, D., J. Pan, and K. Singleton. 2000. Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica* 68: 1343–1376.

Eraker, B. 2004. Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *The Journal of Finance* 59: 1367–1403.

Figueroa-López, J. E., and C. Mancini. 2019. Optimum Thresholding Using Mean and Conditional Mean Squared Error. *Journal of Econometrics* 208: 179–210.

Foschi, R., F. Lilla, and C. Mancini. 2019. Warnings about Future Jumps: Properties of the Exponential Hawkes Model. *Available at SSRN 3459443*.

Geweke, J. F., et al. 1991. "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." Technical report, Federal Reserve Bank of Minneapolis.

Gresnigt, F., E. Kole, and P. H. Franses. 2016. Specification Testing in Hawkes Models. *Journal of Financial Econometrics* 15: 139–171.

Grossman, S. 1976. On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information. *The Journal of Finance* 31: 573–585.

Hawkes, A. G. 1971a. Point Spectra of Some Mutually Exciting Point Processes. *Journal of the Royal Statistical Society: Series B (Methodological)* 33: 438–443.

Hawkes, A. G. 1971b. Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika* 58: 83–90.

Hess, G. D., and S. Iwata. 1997. Measuring and Comparing Business-Cycle Features. *Journal of Business & Economic Statistics* 15: 432–444.

Jacod, J., V. Todorov, et al. 2010. Do Price and Volatility Jump Together? *The Annals of Applied Probability* 20: 1425–1469.

Kass, R. E., and A. E. Raftery. 1995. Bayes Factors. *Journal of the American Statistical Association* 90: 773–795.

Lee, K., and B. K. Seo. 2017. Marked Hawkes Process Modeling of Price Dynamics and Volatility Estimation. *Journal of Empirical Finance* 40: 174–200.

Lee, K., and B. K. Seo. 2022. Modeling Bid and Ask Price Dynamics with an Extended Hawkes Process and Its Empirical Applications for High-Frequency Stock Market Data. *Journal of Financial Econometrics* 1: 1–14.

Lee, S. S. 2012. Jumps and Information Flow in Financial Markets. *The Review of Financial Studies* 25: 439–479.

Liniger, T. J. 2009. Multivariate Hawkes Processes. PhD thesis, ETH Zurich.

Mancini, C., V. Mattiussi, and R. Renò. 2015. Spot Volatility Estimation Using Delta Sequences. *Finance and Stochastics* 19: 261–293.

Maneesoonthorn, W., C. S. Forbes, and G. M. Martin. 2017. Inference on Self-Exciting Jumps in Prices and Volatility Using High-Frequency Measures. *Journal of Applied Econometrics* 32: 504–532.

Merton, R. C. 1976. Option Pricing When Underlying Stock Returns Are Discontinuous. *Journal of Financial Economics* 3: 125–144.

Nakajima, J., and Y. Omori. 2009. Leverage, Heavy-Tails and Correlated Jumps in Stochastic Volatility Models. *Computational Statistics & Data Analysis* 53: 2335–2353.

Raftery, A. E., and S. Lewis. 1991. "How Many Iterations in the Gibbs Sampler?" Technical report, Washington Univ Seattle Dept of Statistics.

Rasmussen, J. G. 2013. Bayesian Inference for Hawkes Processes. *Methodology and Computing in Applied Probability* 15: 623–642.

Rousseeuw, P. J., and A. M. Leroy. 1988. A Robust Scale Estimator Based on the Shortest Half. *Statistica Neerlandica* 42: 103–116.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64: 583–639.