Modeling Financial Contagion Using Mutually Exciting Jump Processes*

Yacine Aït-Sahalia
Department of Economics
Bendheim Center for Finance
Princeton University
and NBER

Julio Cacho-Diaz
Department of Economics
Bendheim Center for Finance
Princeton University

Roger J.A. Laeven
Dept. of Econometrics and OR
Tilburg University
and CentER

This Version: November 25, 2011

Abstract

We propose a model designed to capture the dynamics of asset returns, with periods of crises that are characterized by contagion from one market to another. In the model, a jump in one region of the world increases the intensity of jumps both in the same region (self-excitation) as well as in other regions (cross-excitation), generating episodes of highly clustered jumps across world markets that mimic the features of the real data. We develop and implement estimation and testing procedures for this model, and show that the model fits the data well. The estimates provide evidence of self-excitation both in the US and the other world markets, and of asymmetric cross-excitation, with the US market having more influence on the jump intensity of other markets than the reverse.

Keywords: Jumps; Contagion; Crisis; Mutually exciting processes; Hawkes process.

*We are grateful to seminar and conference participants at the Chinese University in Hong Kong, City University of London, Columbia, Cornell, HEC Paris, HKUST, LSE, Nanyang Technological University, National University of Singapore, Princeton, Seoul National University, Singapore Management University, Tilburg, Toulouse, University of Amsterdam, University of Melbourne, University of Technology Sydney, the AFMATH Conference in Brussels, the Cambridge-Princeton Conference, the Credit Risk Conference in Venice, the Econometric World Congress in Shanghai, the TCF Workshop on Lessons from the Credit Crisis, the SoFiE Conference in Melbourne, the WONDER Research Afternoon in Tilburg, and in particular to Kenneth Lindsay, for their comments and suggestions. This research was funded in part by the NSF under grant SES-0850533 (Aït-Sahalia) and by the NWO under grants Veni-2006 and Vidi-2009 (Laeven). *Matlab* code to implement the estimation procedure developed in this paper is available from the authors upon request. Email addresses: yacine@princeton.edu; jcacho@princeton.edu; R.J.A.Laeven@uvt.nl.

1. Introduction

Large drops in asset markets are very unlikely to occur under standard Brownian-driven statistical models, at least with volatility variables calibrated to realistic values. Even more unlikely would be crashes that happen in not just one, but multiple markets around the world at nearly the same time. And, even more unlikely would be further large price moves that happen in close succession over the following days, like earthquake aftershocks. Yet, despite the predictions of standard models, recurring crises happen every decade or so.

Of course, jump processes can be employed to capture the large moves in asset markets that continuous models are unable to generate by themselves. But the interplay between the various jump terms across markets and over time is complex, and standard jump specifications appear unable to replicate those patterns. In fact, what characterizes a crisis is typically not the initial jump, but the amplification that takes place subsequently over hours or days, and the fact that other markets become affected as well. Figures 1 and 2 illustrate two recent examples of such patterns, which took place in February 2007 and October 2008, respectively. Figure 1 describes the aftermath of a sharp decline that originated in the Chinese stock market. Even though this drop was perhaps not the only information that caused the European and US markets to fall,[1] its effect was felt in other markets in the form of a major downfall repeated over days that by far exceeded the typical market reaction to macroeconomic news releases. In another example, Figure 2 shows what happened on and in the few days that followed October 3, 2008, when the prospects of a $700 billion bailout of the US financial sector were assessed in markets around the world. In a typical jump model with arrival rates calibrated to historical data, where jumps are inherently rare, occurring perhaps once a year on average, observing such patterns of multiple jumps in close succession over hours or days and across different markets would be very unlikely.

More generally, when observed over a longer time period, as in Figure 3, where we plot all the large daily returns in six world regions' stock markets over the last thirty years, it is apparent that large moves tend to appear in clusters, both in time (horizontally) and in space (vertically, across markets).[2] Both effects have been documented separately in different literatures: jump clusters over

[1] There were some concomitant macroeconomic news on durable goods in the US, as well as an announcement by Freddie Mac of tightened lending standards and suspension of purchases of the riskiest subprime mortgages.

[2] Jump clustering in time is a strong effect in the data. For example, from mid-September to mid-November 2008, the US stock market jumped by more than 5% on eight separate days. In the post-World War II era, jumps of magnitude greater than 5% of either sign have only happened on sixteen other days. And intra-day fluctuations were even more pronounced: during the same two months, the range of intraday returns exceeded 10% during fourteen days. Cross-sectionally, the example in Figure 1 is one of the few cases in which the US stock market is affected by a non-US stock market jump. Usually, stock markets around the world appear to take their cues from the US market. A more typical situation is therefore that illustrated in Figure 2 where US news and market events lead successive market moves elsewhere around the world. As becomes apparent from inspection of the figures, the shocks do not

time have generally been identified in the financial econometrics literature,[3] while clusters in space have been the focus of the contagion literature in international finance.[4]

Our objective in this paper is to model this type of financial propagation. For this, we need to leave the widely applied class of Lévy jumps, such as the compound Poisson process, which is the usual class of driving processes employed in the literature. Lévy processes have independent increments: as a result, they do not allow for any type of serial dependence, whence propagation of jumps over time as well as propagation of jumps across markets are key components we wish to capture. So we employ a different model, based on mutually exciting jump processes, known as Hawkes processes. In a Hawkes process, a jump in one market raises the probability of future jumps both in the same market and elsewhere, thereby generating returns that look just like Figures 1 and 2. Jumps in asset returns "self-excite" both in space (the other markets) and in time (the future trading sessions in each market).[5]

In order for the asset returns process to be stationary, we make the jump intensities, which increase with each jump, mean revert until the next jump. Since the jumps' intensities rise in response to the jumps themselves, a feedback element is introduced. Combined with mean-reversion, this aspect of the model plays the same role for jumps as ARCH does for volatility (see Engle (1982)). Namely, the ARCH model introduces feedback from returns to volatility and back: large returns lead to large volatilities which then make it more likely to observe large returns. In the absence of further excitation, volatility then reverts to its steady state level. Here, similarly, jumps lead to larger jump intensities, which then make it more likely to observe further jumps. In the absence of further excitation, jump intensities then revert to their steady state level.

The paper makes two contributions. First, compared to the existing use of Hawkes processes as pure jump processes, we propose a model consisting of a mutually exciting jump component added

occur simultaneously, especially when viewed at intradaily frequencies. It takes some time for the transmission, if at all, to take place. Differences in time zones and trading hours require careful consideration; our treatment of the data is discussed later in the paper.

[3] In a univariate time series setting, Maheu and McCurdy (2004), Yu (2004) and Lee and Mykland (2008) provide evidence for jump clustering.

[4] See Ehrmann, Fratzscher, and Rigobon (2010) and the review of Dungey and Gonzalez-Hermosillo (2005).

[5] Hawkes processes were originally proposed as mathematical models to represent the transmission of contagious diseases in epidemiology. They have also found applications in neurophysiology and in the modeling of earthquake occurrences (see, e.g., Brillinger (1988) and Ogata and Akaike (1982)), and in genome analysis (see Reynaud-Bouret and Schbath (2010)). In market microstructure, Bowsher (2007) employs them to jointly model transaction times and price changes at high frequency. Salmon and Tham (2007) study how trading activity at one maturity of the yield curve affects and is affected by trading at other maturities. Self-exciting models have recently been employed to model joint defaults in portfolios of credit derivatives (see Errais, Giesecke, and Goldberg (2010).) Hawkes processes have also been proposed in the literature on social interactions to model the "viral propagation" of some phenomena, such as the viewing of YouTube videos (see Crane and Sornette (2008)), and to model the occurrence of certain types of crime, such as burglaries, which are empirically highly clustered in time and space (see Mohler, Short, Brantingham, Schoenberg, and Tita (2011).)

to a continuous Brownian component with stochastic volatility, as well as a drift term, which we refer to as a *Hawkes jump-diffusion* model by analogy with the Poissonian jump-diffusion model familiar to financial economists since Merton (1976). By contrast, in the pure jump Hawkes models that have been previously employed, sample paths are piecewise constant, which is appropriate to model point events such as financial default or earthquake occurrences, but not for asset prices: we also need a Brownian component, or alternatively, an infinite activity jump component to capture non-crisis periods. Second, we develop and implement estimation and testing procedures for this model, which are new in the context of Hawkes processes. The estimation procedure is based on the generalized method of moments (GMM). The technical difficulty involves integrating out the latent state variables (stochastic volatilities and intensities) of the model when computing the moment functions, or estimating equations, which we will provide in closed-form.

We estimate the model using international stock index returns for six world regions. The empirical analysis indicates that both the US market as well as other markets strongly self-excite over time. Cross-sectionally, we find evidence of asymmetry: estimates from the model imply that US jumps tend to get reflected quickly in most other markets, while statistical evidence for the reverse transmission is much less pronounced. We further show that a model based on Poissonian (hence non-mutually exciting, non-contagion) jumps is empirically rejected against the mutually exciting model based on Hawkes jumps, which allows for contagion.

The rest of this paper is organized as follows: In Section 2, we present the model of asset returns, and discuss some of its properties. In Section 3, we develop our estimation procedure. Section 4 contains the empirical analysis, where we show that the model fits the data well and test the model against alternatives including a Poissonian jump-diffusion model. Conclusions are in Section 5.

2. Asset Return Dynamics

2.1 Jump Dynamics: Mutually Exciting Processes

Hawkes (1971b) (see also Hawkes (1971a), Hawkes and Oakes (1974) and Oakes (1975)) introduced mutually exciting processes, which are special cases of path-dependent point processes. The intensities of a multivariate mutually exciting process depend on the paths of the point process; hence, the jump intensities are themselves stochastic processes and will be part of the state vector. The couple consisting of the jump process and its intensity remains a Markov process. The jump intensities ramp up in response to jumps in one of the marginal point processes. We consider m point processes $N_{i,t}$, $i = 1, \ldots, m$, one for each of the m regions of the world.

Similar to the familiar Poisson process, the Hawkes process is defined by its intensity process $\lambda_{i,t}$

which describes the $\mathcal{F}_t-$conditional mean jump rate per unit of time, namely

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

but instead of being constant the jump intensities follow the dynamics

$$\lambda_{i,t} = \lambda_{i,\infty} + \sum_{j=1}^{m} \int_{-\infty}^{t} g_{i,j}\left(t-s\right) dN_{j,s}, \qquad i = 1, \ldots, m. \tag{2.2}$$

In other words, each previous jump $dN_{j,s}$, $s \in (-\infty, t)$, $j = 1, \ldots, m$, raises the jump intensities $\lambda_{i,t}$. Taken jointly, (N, λ) is a Markov process. The distribution of the jump processes $N_{j,s}$ is determined by that of the intensities $\lambda_{j,s}$. Each compensated process $N_{i,t} - \int_{-\infty}^{t} \lambda_{i,s} ds$ is a local martingale. Hawkes processes are related to, but different from, doubly stochastic Poisson (or Cox) processes. In a doubly stochastic Poisson process, the jump intensities are also stationary stochastic processes but the conditioning σ-algebra in (2.1) is $\{N_{i,s}(s \le t) : \lambda_{i,s}(-\infty < s < \infty)\}$, i.e., the path of $\lambda_{i,t}$ is not affected by the path of $N_{i,t}$.[6]

We assume that in (2.2) the constant parameters $\lambda_{i,\infty} \ge 0$ for all $i = 1, \ldots, m$, that the real-valued functions $g_{i,j}(u) \ge 0$ for all $u \ge 0$ and for all $i, j = 1, \ldots, m$. Notice that $\lambda_{i,\infty}, g_{i,j}$ being non-negative for all $i, j = 1, \ldots, m$ ensures the non-negativity of the intensity processes with probability one. Let Λ_∞ denote the $m \times 1$ vector with components $\lambda_{i,\infty}$ and $\Gamma := \int_0^\infty G_u du$ the $m \times m$ matrix where G_u is the matrix with elements $g_{i,j}(u)$'s. Let $\lambda_i := \mathbb{E}\left[\lambda_{i,t}\right]$ denote the unconditional expected jump intensity, and Λ the $m \times 1$ vector with components λ_i. Since from (2.1), $\mathbb{E}\left[dN_{i,s}\right] = \lambda_i ds$, we see that

$$\lambda_i = \lambda_{i,\infty} + \sum_{j=1}^{m} \lambda_j \int_{-\infty}^{t} g_{i,j}\left(t-s\right) ds = \lambda_{i,\infty} + \sum_{j=1}^{m} \left(\int_0^\infty g_{i,j}\left(u\right) du \right) \lambda_j, \tag{2.3}$$

or in vector form $\Lambda = \Lambda_\infty + \Gamma \cdot \Lambda$. Therefore $\Lambda = \left(I - \Gamma\right)^{-1} \Lambda_\infty$, where I is the identity matrix, and we assume that all the elements of Λ are positive and finite. This ensures stationarity of the model.

[6] Alternative models exist with state-dependent jump intensities, generating time-variation in the arrival rate of jumps; see, e.g., Bates (2000), Pan (2002), Eraker (2004) and Johannes (2004). In these models, jump intensities depend on contemporaneous other variables, often the asset's volatility, but not on the past jumps of the price process: they are therefore different from the feedback element that is our focus.

2.2 Full Dynamics: Mutually Exciting Jump-Diffusion

The mutually exciting jump process constitutes only part of our model of asset returns. We assume that asset log-returns follow the semimartingale dynamics

$$dX_{i,t} = \mu_i dt + \sigma_i dW_{i,t} + Z_{i,t} dN_{i,t}, \qquad i = 1, \ldots, m, \tag{2.4}$$

which consists of a drift term, a volatility term, and mutually exciting jumps.

$W_t := [W_{1,t}, \ldots, W_{m,t}]'$ is an m-dimensional vector of standard Brownian motions with constant correlation coefficients $\rho_{i,j},\ i,j = 1, ..., m,\ Z_t := [Z_{1,t}, \ldots, Z_{m,t}]'$ is the vector of jump sizes, cross-sectionally and serially independently distributed with laws F_{Z_i} supported on $(-\infty, \infty)$, and $N_t := [N_{1,t}, \ldots, N_{m,t}]'$ is the vector of Hawkes processes just described. Z_i may be replaced by $\log\left(1+\tilde{Z}_i\right)$ so that the discontinuous part in the differential of the stochastic exponential $e^{X_{i,t}}$ becomes $\tilde{Z}_i e^{X_{i,t}} dN_{i,t}$. For notational convenience we write Z_i rather than $\log\left(1+\tilde{Z}_i\right)$. We will leave the distribution of jump sizes essentially unrestricted, so asymmetries such as negative jumps (crashes) being more likely than positive jumps (booms), can be built into the jump size distributions.

In the base model (2.4), the quantities μ_i and σ_i are constant parameters. In this case we assume that Σ, the $m \times m$-dimensional variance-covariance matrix of the risk coming from the continuous part of the model, with elements $\Sigma_{i,j} = \rho_{i,j}\sigma_i\sigma_j$, is a non-singular matrix. We always assume that the vector of Brownian motions W, the vector of jump sizes Z and the vector of jump processes N are mutually independent. In an extension of the model (2.4), we allow for stochastic volatility:

$$dX_{i,t} = \mu_i dt + \sqrt{V_{i,t}} dW_{i,t}^X + Z_{i,t} dN_{i,t}, \tag{2.5}$$

where the instantaneous variance follows the Heston (1993) model

$$dV_{i,t} = \kappa_i(\theta_i - V_{i,t})dt + \eta_i \sqrt{V_{i,t}} dW_{i,t}^V, \tag{2.6}$$

with κ_i, θ_i, η_i constant parameters. Model (2.4) corresponds to the special case where the initial value is $V_{i,0} = \theta_i$ and $\eta_i = 0$. $V_{i,t}$ follows a square root process and is bounded below by zero. The boundary value zero cannot be achieved as long as Feller's condition, $2\kappa_i\theta_i \geq \eta_i^2$, is satisfied.

The extended model allows for correlations among the individual Brownian motions driving equations (2.6) as well as between the individual Brownian motions driving equations (2.5) and (2.6), thereby capturing a potential leverage effect. But in the presence of systematic jumps, the Brownian correlation, even if it increases, will not play a major role: in times of crisis, jumps swamp everything else.

While the estimation procedure can cover general stochastic volatility specifications, we need to settle on a specific model with interpretable parameters to do the actual parametric estimation, and for this purpose we assume that the volatility follows the model (2.6). *Hawkes jump-diffusion* processes, as we refer to the mutually exciting jump-diffusion, will generate observed time-varying correlations and maximal correlations around crisis times, due to the systematic jumps. Isolating a change in the structure of the Brownian variance-covariance matrix Σ becomes difficult in this context, because linear correlation measures weigh equally all returns; as a result, they will tend to average out any contagion that happens over a limited number of days among the comparatively large number of days where no jumps occur. Similarly, the observed clustering of extreme returns is generated by Hawkes jump-diffusion processes in a different manner from pure stochastic volatility models where periods of high volatility can also lead to clusters of larger absolute returns, although not of a magnitude and rate of occurrence compatible with what is actually observed in the data.

2.3 Mean Reversion in Jump Intensities

The tractability of the jump part of our model, and hence the possibility of estimating it, depends on the parameterization of the intensity processes $\lambda_{i,t}$. A special case of interest occurs when

$$g_{i,j}(t-s) = \beta_{i,j} e^{-\alpha_i (t-s)}, \qquad s < t, \quad i,j = 1,\ldots,m, \tag{2.7}$$

with $\alpha_i > 0$, $\beta_{i,j} \geq 0$ for all $i,j = 1,\ldots,m$. In this case, a jump in asset prices causes the intensities to jump up, and then the intensity decays exponentially back: $\lambda_{i,t}$ jumps by $\beta_{i,j}$ whenever a shock in sector j occurs, and then decays back towards a level $\lambda_{i,\infty}$ at speed α_i. Asset return dynamics of equity markets that are highly interrelated may exhibit jumps that take place almost simultaneously. This corresponds to the case in which mutual excitation is very strong (large $\beta_{i,j}$, $i \neq j$).

With this model, the Γ matrix is given by $\Gamma = [\beta_{i,j}/\alpha_i]_{i,j=1,..,m}$. Under exponential decay (2.7), each jump intensity has the mean-reverting dynamics

$$d\lambda_{i,t} = \alpha_i \left(\lambda_{i,\infty} - \lambda_{i,t}\right) dt + \sum_{j=1}^{m} \beta_{i,j} dN_{j,t}. \tag{2.8}$$

In more general specifications, the intensity may depend not only on the amount of time elapsed since a jump event but also on the size of past jump events.

The jump part of the model is able to generate the two features we are after: First, jump activity is variable over time, with many (of the few) jumps typically being concentrated in short periods of time; this is jump clustering. Second, adverse shocks propagate across the world in a contagious way, with adverse events in one region of the world seemingly increasing the likelihood of shocks in other regions of the world; this is jump propagation. Under exponential decay, the element-by-element Laplace transform $V_N^*(s)$ of the covariance density matrix $V_N(\tau)$ is given by

$$V_N^*(s) = (I - G^*(s))^{-1} \left(G^*(s)\mathrm{Diag}(\Lambda) + \beta V_N^* \,'\{\alpha\}/\{\alpha + s\} \right), \tag{2.9}$$

with $G(\tau)$ the $m \times m$-matrix given by $G(\tau) = [\beta_{i,j} e^{-\alpha_i \tau}]_{i,j=1,\ldots,m}$ and $G^*(s)$ its element-by-element Laplace transform. The indices of $\{a\}$ and $\{a + s\}$ correspond to the element-by-element indices of β.

2.4 Jump Size Distribution

We will provide expressions for the moments of the process as functions of the generic moments of the jump size Z_t, which we denote $M[Z, k] := \mathbb{E}\left[Z_t^k\right]$. In order to estimate a specific parametric model, however, we need to parameterize these moments to reduce the dimensionality of the parameter space, given that the distribution of jumps of finite activity is difficult to pin down since large jumps are by nature rare. For this purpose, we will assume that Z_i is a scalar random variable with cumulative probability distribution

$$F_{Z_i}(x) = \begin{cases} p_i e^{-\gamma_{i,1}(-x)}, & -\infty < x \leq 0; \\ p_i + (1 - p_i)(1 - e^{-\gamma_{i,2} x}), & 0 < x < \infty; \end{cases} \tag{2.10}$$

where $\gamma_{i,1}, \gamma_{i,2} > 0$ and $0 \leq p_i \leq 1$, $i = 1, \ldots, m$ so that

$$\mathbb{E}\left[Z_i^k\right] = (-1)^k \frac{k!\ p_i}{\gamma_{i,1}^k} + \frac{k!\ (1 - p_i)}{\gamma_{i,2}^k}, \qquad k = 1, 2, \ldots. \tag{2.11}$$

In our empirical study below we use equity index data. As such, our analysis does not suffer from survivorship bias. However, survivorship bias, when present (for example, when individual stock returns are used), can be dealt with in this model. Explicitly modeling the survival and death process involves introducing a point mass at $Z_i = -\infty$ in the distribution of the jump size for log-returns.

3. Estimation Procedure

We will develop a GMM-based estimation procedure for Hawkes-driven processes. Neither the point processes $N_{i,t}$ and the intensity processes $\lambda_{i,t}$, nor the stochastic volatilities $V_{i,t}$, $i = 1, \ldots, m$, are directly observable. Instead, what is observed are the log-returns $X_{i,t}$. It turns out that we can derive explicit expressions for the moments of the log-returns that are implied by the model, as a function only of the observable state variables, in effect integrating out the unobservable state variables. We achieve this by first computing the conditional moments using the *full state vector*: asset returns, stochastic volatilities and jump intensities; then conditioning down by taking expected values over the *latent state variables*: volatilities and jump intensities. This results in expressions for the moments we derive that depend only upon the *observable state variables*, namely the asset returns, but involve all the parameters of the model, including those driving the latent state variables.

3.1 Explicit Expressions for the Moments: Markov Infinitesimal Generator

The key to a tractable GMM estimation procedure is the availability in closed-form of the moment functions. Having the moment functions in closed-form means that the effort involved in minimizing the GMM criterion function becomes as minimal and straightforward as can be, given the inherent difficulty associated with the dimensionality of the model and the number of parameters. The moment conditions that we will use to carry out GMM estimation of the model are:

$$\begin{cases} \mathbb{E}\left[\Delta X_{i,t}\right] \\ \mathbb{E}\left[\left(\Delta X_{i,t} - \mathbb{E}\left[\Delta X_{i,t}\right]\right)^r\right], \qquad r = 2, \ldots, 4 \\ \mathbb{E}\left[\Delta X_{i,t}\Delta X_{j,t} - \mathbb{E}\left[\Delta X_{i,t}\right]\mathbb{E}\left[\Delta X_{j,t}\right]\right], \qquad i \neq j \\ \mathbb{E}\left[\Delta X_{i,t+\tau}\Delta X_{j,t} - \mathbb{E}\left[\Delta X_{i,t}\right]\mathbb{E}\left[\Delta X_{j,t}\right]\right], \qquad \tau > 0. \end{cases} \tag{3.1}$$

As we will see, the reason for using these moment functions is two-fold. First, they are "natural" in the sense of being easily interpretable: variance, skewness, kurtosis, autocovariances. Second, and more importantly, each moment function plays a specific role in identifying parts of the model. This is particularly important when estimating higher-dimensional models, since this feature means that we can decompose a problem with many parameters into many smaller-dimensional ones, thereby facilitating the estimation in practice.

To compute the unconditional moments, we first exploit the Markovian nature of the model to compute the one-step ahead conditional moments using the full state vector: asset returns, stochastic volatilities and jump intensities. Next, we take expected values, integrating out the latent state variables: volatilities and jump intensities. Doing so, we obtain expressions that depend only upon the observable state variables: asset returns. These expressions, however, contain all the parameters of the model, including those related to the latent state variables.

To compute the conditional moments, we use the explicit expression of the generator of the Markov process (2.4) or (2.5)-(2.6) driven by the processes (W, N, λ). Specifically, the computation of the moment functions reduces to the evaluations of conditional expectations of functions of the form $f(y_1, y_0, \delta)$ where y_1 and y_0 denote log-returns separated by some function of the sampling interval δ. We assume that sampling is equidistant in time. We need to evaluate expressions of the form $\mathbb{E}_{Y_1,Y_0}[f(Y_1, Y_0, \Delta)]$. The dynamics of the system depends upon additional latent variables which at the same instants we denote by $\xi_1 = (v_1, l_1)$ and $\xi_0 = (v_0, l_0)$, respectively, where v is the volatility and l the jump intensity. The standard infinitesimal generator $\mathcal{A}$ is the operator which, when applied to a function $g(y_1, y_0, \xi_1, \xi_0, \delta)$ in its domain, returns the function $\mathcal{A} \cdot g$ given in full generality by

$$\begin{aligned}\mathcal{A} \cdot g = {} & \frac{\partial g}{\partial \delta} + \sum_{i=1}^{m} \mu_i^X \frac{\partial g}{\partial y_{1,i}} + \sum_{i=1}^{m} \mu_i^V \frac{\partial g}{\partial v_{1,i}} + \sum_{i=1}^{m} \mu_i^\lambda \frac{\partial g}{\partial l_{1,i}} \\ & + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_{ij}^{YY} \frac{\partial^2 g}{\partial y_{1,i} \partial y_{1,j}} + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_{ij}^{YV} \frac{\partial^2 g}{\partial y_{1,i} \partial v_{1,j}} + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_{ij}^{VV} \frac{\partial^2 g}{\partial v_{1,i} \partial v_{1,j}} \\ & + \sum_{i=1}^{m} l_{1,i} \int_{z_i} \left\{ g(y_1 + z_i, y_0, v_1, l_1 + \beta_i, v_0, l_0, \delta) - g(y_1, y_0, v_1, l_1, v_0, l_0, \delta) \right\} f_{Z_i}(z_i) dz_i,\end{aligned} \tag{3.2}$$

where μ^X is the vector of expected log-returns, μ^V the vector of stochastic volatility drifts, μ^λ the vector of jump intensity drift, σ^{YY} the variance-covariance matrix of log-returns, σ^{YV} the variance-covariance matrix of interactions between returns and volatilities, σ^{VV} the variance-covariance matrix of stochastic volatilities and finally $\beta_i = [\beta_{i,1}, ..., \beta_{i,m}]'$ the vector of excitation parameters for the i^{th} jump term. In the case of the mean-reverting jump intensity model described in Section 2.3, we have $\mu_i^\lambda = \alpha_i (\lambda_{i,\infty} - l_{1,i})$.

The usefulness of the infinitesimal generator for our purpose lies in the fact that

$$\begin{aligned}\mathbb{E}_{Y_1,\xi_1}[g(Y_1, Y_0, \xi_1, \xi_0, \Delta) | Y_0, \xi_0] &= \exp(\Delta \mathcal{A}) \cdot g(Y_0, Y_0, \xi_0, \xi_0, 0) \\ &= \sum_{j=0}^{J} \frac{\Delta^j}{j!} \left(\mathcal{A}^j \cdot g \right) (Y_0, Y_0, \xi_0, \xi_0, 0) + O_p\left(\Delta^{J+1} \right),\end{aligned} \tag{3.3}$$

where subscripts in $\mathbb{E}_{Y_1,\xi_1}$ indicate the random variables that the expected value operates on, and $\mathcal{A}^j \cdot g$ is defined recursively by $\mathcal{A}^j \cdot g = \mathcal{A} \cdot (\mathcal{A}^{j-1} \cdot g)$ for all $j \geq 1$. As we will see in the theorems that follow, (3.2) and its iterates $\mathcal{A}^j \cdot g$, hence the terms in (3.3), can be evaluated in closed-form for the moment functions of interest given in (3.1). While the starting moment function $f(y_1, y_0, \delta)$ does not depend directly upon the latent variables (ξ_1, ξ_0), $\mathcal{A} \cdot f$ and successive iterates will, since the coefficients $\{\mu^X, \mu^V, \mu^\lambda, \sigma^{YY}, \sigma^{YV}, \sigma^{VV}, \beta_i\}$ will in general depend on the full state variables $[X, V, \lambda]$.

In other words, we can obtain the conditional expectation of g, using the full state vector including its unobservable components, $\mathbb{E}_{Y_1,\xi_1}\left[g(Y_1, Y_0, \xi_1, \xi_0, \Delta)|Y_0, \xi_0\right]$. We then need to "condition down" by integrating out the unobservable state variables in order to produce moment functions that can be fitted to the log-returns data. All the expectations are taken with respect to the law of the process at the true parameter values. From the law of iterated expectations, we have that

$$\begin{aligned}\mathbb{E}_{Y_1,Y_0,\xi_1,\xi_0}\left[g(Y_1, Y_0, \xi_1, \xi_0, \Delta)\right] &= \mathbb{E}_{Y_1,Y_0,\xi_1,\xi_0}\left[\mathbb{E}_{Y_1,\xi_1}\left[g(Y_1, Y_0, \xi_1, \xi_0, \Delta)|Y_0, \xi_0\right]\right] \\ &= \sum_{j=0}^{J} \frac{\Delta^j}{j!} \mathbb{E}_{Y_0,\xi_0}\left[\left(\mathcal{A}^j \cdot g\right)(Y_0, Y_0, \xi_0, \xi_0, 0)\right] + O_p\left(\Delta^{J+1}\right), \qquad (3.4)\end{aligned}$$

so the last step in the necessary calculations will involve computing unconditional expectations with respect to the stationary law of the state variables.

3.2 The Univariate Model

For ease of exposition, we first provide the expressions of these moments in the univariate case $m = 1$. In this situation, there is no cross-sectional excitation, only time-series excitation: we are considering a model for a single asset with Heston-type stochastic volatility and jumps that self-excite, meaning that future jump intensities depend upon the history of their own past jumps:

$$\begin{cases} dX_t = \mu dt + \sqrt{V_t} dW_t^X + Z_t dN_t \\ dV_t = \kappa(\theta - V_t)dt + \eta\sqrt{V_t} dW_t^V \\ d\lambda_t = \alpha\left(\lambda_\infty - \lambda_t\right) dt + \beta dN_t \end{cases} \qquad (3.5)$$

with $\mathbb{E}\left[dW_t^X dW_t^V\right] =: \rho^V dt$ and $\lambda := \mathbb{E}\left[\lambda_t\right] = \alpha\lambda_\infty/(\alpha - \beta)$. Note that classical compound Poisson process jumps are obtained when $\beta = 0$ and $\lambda_0 = \lambda_\infty$. Then $\lambda_t = \lambda_\infty = \lambda$ at all t.

For simplicity, we make the jump intensities respond only to jumps, not to their sizes or signs. This is to preserve the autonomous nature of the couple (N, λ) which facilitates the analysis of the model. But it is possible to introduce a level effect and an asymmetry effect by replacing the term βdN_t in the third equation of (3.5) with a term $\phi(Z_t)dN_t$ where ϕ is a function selected jointly with the distribution of Z_t to maintain the positivity and stationarity of λ_t. Such an effect could for instance capture a larger increase in jump intensity following a large negative jump in the asset price.

In the simple model, we can leave the distribution of the jump size essentially unrestricted, and provide expressions as functions of the moments of the jump size Z_t, for which we write $M[Z,k] := \mathbb{E}\left[Z_t^k\right]$. Our first main result gives the moment functions explicitly:

Theorem 1. *For the univariate model (3.5), the moments are given in closed-form up to order Δ^2 by the following expressions*

$$\begin{aligned}
\mathbb{E}\left[\Delta X_t\right] &= (\mu + \lambda M\left[Z,1\right])\Delta + o(\Delta^2) \\
\mathbb{E}\left[(\Delta X_t - \mathbb{E}\left[\Delta X_t\right])^2\right] &= (\theta + \lambda M\left[Z,2\right])\Delta + \frac{\beta\lambda\left(2\alpha-\beta\right)}{2(\alpha-\beta)} M\left[Z,1\right]^2 \Delta^2 + o(\Delta^2) \\
\mathbb{E}\left[(\Delta X_t - \mathbb{E}\left[\Delta X_t\right])^3\right] &= \lambda M\left[Z,3\right]\Delta \\
&\quad + \frac{3}{2}\left(\eta\theta\rho^V + \frac{(2\alpha-\beta)\beta\lambda M\left[Z,1\right] M\left[Z,2\right]}{(\alpha-\beta)}\right)\Delta^2 + o(\Delta^2) \\
\mathbb{E}\left[(\Delta X_t - \mathbb{E}\left[\Delta X_t\right])^4\right] &= \lambda M\left[Z,4\right]\Delta + \left(\frac{3\theta\eta^2}{2\kappa} + 3\theta^2 + 6\theta\lambda M\left[Z,2\right]\right. \\
&\quad + 3\lambda\left(\lambda + \frac{(2\alpha-\beta)\beta}{2(\alpha-\beta)}\right) M\left[Z,2\right]^2 \\
&\quad \left. + \frac{2(2\alpha-\beta)\beta\lambda M\left[Z,1\right] M\left[Z,3\right]}{(\alpha-\beta)}\right)\Delta^2 + o(\Delta^2)
\end{aligned}$$

while the autocorrelation function of the process is given by

$$\mathbb{E}\left[(\Delta X_t - \mathbb{E}\left[\Delta X_t\right])(\Delta X_{t+\tau} - \mathbb{E}\left[\Delta X_{t+\tau}\right])\right] = \frac{\beta\lambda(2\alpha-\beta)}{2(\alpha-\beta)} e^{-(\alpha-\beta)\tau} M\left[Z,1\right]^2 \Delta^2 + o(\Delta^2)$$

for all $\tau > 0$.

Intuitively, the identification of the parameters is achieved as follows: the higher order moments (3 and 4) isolate the jump parameters at the leading order, while the variance puts them on an equal footing with the diffusive parameters.[7] The autocovariance isolates the jump parameters. Indeed, if the model had no jump component, then from (3.5) with $Z_t \equiv 0$, the law of iterated expectations implies that

$$\begin{aligned}
\mathbb{E}\left[(\Delta X_t)(\Delta X_{t+\tau})\right] &= \mathbb{E}\left[(\Delta X_t)\,\mathbb{E}\left[(\Delta X_{t+\tau})\,|\mathcal{F}_{t+\tau}\right]\right] = \mathbb{E}\left[(\Delta X_t)(\mu\Delta)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[(\Delta X_t)\,|\mathcal{F}_t\right](\mu\Delta)\right] = \mu^2\Delta^2
\end{aligned}$$

[7] We use regular moments as opposed to absolute moments. The use of absolute moments –especially absolute moments of order less than 1 (see Aït-Sahalia (2004))– may be considered in addition, especially for estimating parameters of the volatility process. Since we are mainly interested in the parameters of the jump process we have chosen to consider only regular moments in the interest of simplicity and parsimony.

and so $\mathbb{E}\left[(\Delta X_t - \mathbb{E}\left[\Delta X_t\right])(\Delta X_{t+\tau} - \mathbb{E}\left[\Delta X_{t+\tau}\right])\right] = 0$. Thus any autocovariance is due to the jump component. Further, if the jump component is Poissonian, then the increments would be independent and it is the self-excitation of the jumps that gives rise to the autocorrelation. Thus the observed autocovariance of the increments isolates the self-exciting component of the model. As noted above, the model reduces to a Poissonian jump-diffusion if $\beta = 0$. Then the model no longer generates any autocorrelation in returns.

3.3 The Bivariate Model

We now turn to the bivariate case, $m = 2$. For ease of exposition, we restrict attention here to the more tractable triangular excitation case, where $\beta_{12} = 0$, with state-independent volatilities, and consider the general non-triangular case with state-dependent volatilities in the Appendix. That is, the base model is

$$\begin{cases} dX_{1t} = \mu_1 dt + \sigma_1 dW_t^1 + Z_{1t} dN_{1t} \\ dX_{2t} = \mu_2 dt + \sigma_2 dW_t^2 + Z_{2t} dN_{2t} \\ d\lambda_{1t} = \alpha_1 \left(\lambda_{1\infty} - \lambda_{1t}\right) dt + \beta_{11} dN_{1t} \\ d\lambda_{2t} = \alpha_2 \left(\lambda_{2\infty} - \lambda_{2t}\right) dt + \beta_{21} dN_{1t} + \beta_{22} dN_{2t} \end{cases} \tag{3.6}$$

with $\mathbb{E}[dW_t^1 dW_t^2] =: \rho dt$.

Theorem 2. *For the bivariate model (3.6), the mean and variance of the log-returns are given in closed-form up to order Δ^2 by the following expressions*

$$\mathbb{E}\left[\Delta X_{1t}\right] = (\mu_1 + \lambda_1 M[Z_1, 1])\Delta + o(\Delta^2)$$
$$\mathbb{E}\left[\Delta X_{2t}\right] = (\mu_2 + \lambda_2 M[Z_2, 1])\Delta + o(\Delta^2)$$

and

$$\begin{aligned} \mathbb{E}\left[(\Delta X_{1t} - \mathbb{E}\left[\Delta X_{1t}\right])^2\right] &= \left(\sigma_1^2 + \lambda_1 M[Z_1, 2]\right)\Delta + \tfrac{(2\alpha_1 - \beta_{11})\beta_{11}\lambda_1}{2(\alpha_1 - \beta_{11})} M[Z_1, 1]^2 \Delta^2 + o(\Delta^2) \\ \mathbb{E}\left[(\Delta X_{2t} - \mathbb{E}\left[\Delta X_{2t}\right])^2\right] &= (\sigma_2^2 + \lambda_2 M\left[Z_2, 2\right])\Delta + \left(\tfrac{\beta_{21}^2\left(\alpha_1^2 + (\alpha_1 - \beta_{11})(\alpha_2 - \beta_{22})\right)\lambda_1}{2(\alpha_1 - \beta_{11})(\alpha_2 - \beta_{22})(\alpha_1 + \alpha_2 - \beta_{11} - \beta_{22})}\right. \\ &\quad \left. + \tfrac{(2\alpha_2 - \beta_{22})\beta_{22}\lambda_2}{2(\alpha_2 - \beta_{22})}\right) M[Z_2, 1]^2 \Delta^2 + o\left(\Delta^2\right) \\ \mathbb{E}\left[(\Delta X_{1t} - \mathbb{E}\left[\Delta X_{1t}\right])(\Delta X_{2t} - \mathbb{E}\left[\Delta X_{2t}\right])\right] &= \rho\sigma_1\sigma_2\Delta \\ &\quad + \tfrac{\beta_{21}\left(\alpha_1^2 + (\alpha_1 - \beta_{11})(\alpha_2 - \beta_{22})\right)\lambda_1}{2(\alpha_1 - \beta_{11})(\alpha_1 + \alpha_2 - \beta_{11} - \beta_{22})} M\left[Z_1, 1\right] M\left[Z_2, 1\right] \Delta^2 + o\left(\Delta^2\right). \end{aligned}$$

As expected, the variance places the contributions from the diffusive and jump components of the model on the same level. However, the leading terms of the higher order moments isolate the jump component:

Theorem 3. *The higher order moments of the bivariate model are given up to order Δ^2 by the following expressions:*

$$\begin{aligned}
\mathbb{E}\left[(\Delta X_{1t}-\mathbb{E}\left[\Delta X_{1t}\right])^3\right] &= \lambda_1 M[Z_1,3]\Delta + \tfrac{3(2\alpha_1-\beta_{11})\beta_{11}\lambda_1}{2(\alpha_1-\beta_{11})} M\left[Z_1,1\right] M\left[Z_1,2\right]\Delta^2 + o(\Delta^2) \\
\mathbb{E}\left[(\Delta X_{2t}-\mathbb{E}\left[\Delta X_{2t}\right])^3\right] &= \lambda_2 M[Z_2,3]\Delta + \tfrac{3}{2}\Big(\tfrac{(2\alpha_2-\beta_{22})\beta_{22}\lambda_2}{(\alpha_2-\beta_{22})} \\
&\quad + \tfrac{\left(\alpha_1^2+(\alpha_1-\beta_{11})(\alpha_2-\beta_{22})\right)\lambda_1\beta_{21}^2}{(\alpha_1-\beta_{11})(\alpha_2-\beta_{22})(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})}\Big) M\left[Z_2,1\right] M\left[Z_2,2\right]\Delta^2 + o(\Delta^2) \\
\mathbb{E}\left[(\Delta X_{1t}-\mathbb{E}\left[\Delta X_{1t}\right])^2(\Delta X_{2t}-\mathbb{E}\left[\Delta X_{2t}\right])\right] &= \tfrac{\beta_{21}\left(\alpha_1^2+(\alpha_2-\beta_{22})(\alpha_1-\beta_{11})\right)\lambda_1}{2(\alpha_1-\beta_{11})(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})} \\
&\quad \times M\left[Z_1,2\right] M\left[Z_2,1\right]\Delta^2 + o(\Delta^2) \\
\mathbb{E}\left[(\Delta X_{1t}-\mathbb{E}\left[\Delta X_{1t}\right])(\Delta X_{2t}-\mathbb{E}\left[\Delta X_{2t}\right])^2\right] &= \tfrac{\beta_{21}\left(\alpha_1^2+(\alpha_2-\beta_{22})(\alpha_1-\beta_{11})\right)\lambda_1}{2(\alpha_1-\beta_{11})(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})} \\
&\quad \times M\left[Z_1,1\right] M\left[Z_2,2\right]\Delta^2 + o(\Delta^2)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[(\Delta X_{1t}-\mathbb{E}\left[\Delta X_{1t}\right])^4\right] &= \lambda_1 M[Z_1,4]\Delta + \Big(3\sigma_1^4 + 6M\left[Z_1,2\right]\lambda_1\sigma_1^2 \\
&\quad + \tfrac{3M[Z_1,2]^2\lambda_1(2\alpha_1(\beta_{11}+\lambda_1)-\beta_{11}(\beta_{11}+2\lambda_1))}{2(\alpha_1-\beta_{11})} + \tfrac{2(2\alpha_1-\beta_{11})\beta_{11}\lambda_1}{(\alpha_1-\beta_{11})} M\left[Z_1,1\right] M\left[Z_1,3\right]\Big)\Delta^2 + o(\Delta^2) \\
\mathbb{E}\left[(\Delta X_{2t}-\mathbb{E}\left[\Delta X_{2t}\right])^4\right] &= \lambda_2 M[Z_2,4]\Delta + \Big(3\sigma_1^4 + 6M\left[Z_2,2\right]\lambda_2\sigma_2^2 \\
&\quad + 3M\left[Z_2,2\right]^2(\eta_{22}+\beta_{22}\lambda_2) + 4\left(\eta_{22}+(\beta_{22}-\lambda_2)\lambda_2\right) M\left[Z_2,1\right] M\left[Z_2,3\right]\Big)\Delta^2 + o(\Delta^2).
\end{aligned}$$

The other contemporaneous cross-terms that are not listed above,

$$\mathbb{E}\left[(\Delta X_{1t}-\mathbb{E}\left[\Delta X_{1t}\right])^j(\Delta X_{2t}-\mathbb{E}\left[\Delta X_{2t}\right])^k\right],$$

of order $j+k=4$, are of smaller order $O(\Delta^2)$.

The expressions above depend upon expected values and variances of the jump intensities, which are given by

$$\begin{aligned}
\lambda_1 &:= \mathbb{E}\left[\lambda_{1t}\right] = \tfrac{\alpha_1\lambda_{1,\infty}}{\alpha_1-\beta_{11}} \\
\lambda_2 &:= \mathbb{E}\left[\lambda_{2t}\right] = \tfrac{\alpha_1\beta_{21}\lambda_{1,\infty}+\alpha_1\alpha_2\lambda_{2,\infty}-\alpha_2\beta_{11}\lambda_{2,\infty}}{(\alpha_1-\beta_{11})(\alpha_2-\beta_{22})}
\end{aligned}$$

and

$$\begin{aligned}
\eta_{11} &:= \mathbb{E}\left[\lambda_{1t}^2\right] = \lambda_1^2 + \tfrac{\beta_{11}^2\lambda_1}{2(\alpha_1-\beta_{11})} \\
\eta_{22} &:= \mathbb{E}\left[\lambda_{2t}^2\right] = \lambda_2^2 + \tfrac{\beta_{22}^2\lambda_2}{2(\alpha_2-\beta_{22})} + \tfrac{\left(\alpha_1^2+(\alpha_2-\beta_{22})\alpha_1+\beta_{11}(\beta_{22}-\alpha_2)\right)\lambda_1\beta_{21}^2}{2(\alpha_1-\beta_{11})(\alpha_2-\beta_{22})(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})} \\
\eta_{12} &:= \mathbb{E}\left[\lambda_{1t}\lambda_{2t}\right] = \lambda_2\lambda_1 + \tfrac{(2\alpha_1-\beta_{11})\beta_{11}\beta_{21}\lambda_1}{2(\alpha_1-\beta_{11})(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})}.
\end{aligned}$$

The final result provides the autocovariance function of the bivariate model:

Theorem 4. *For the bivariate model (3.5), the autocovariances of the log-returns at lag $\tau > 0$,*

$$C_{i,j}(\tau) := \mathbb{E}\left[(\Delta X_{j,t}-\mathbb{E}\left[\Delta X_{j,t}\right])(\Delta X_{i,t+\tau}-\mathbb{E}\left[\Delta X_{i,t+\tau}\right])\right],$$

are given in closed-form up to order Δ^2 by the following expressions, which are driven by the mutually *exciting component of the model:*

$$\begin{aligned}
\mathbb{C}_{1,1}(\tau) &= e^{-\tau(\alpha_1-\beta_{11})}a_{11,1}M\left[Z_1,1\right]^2\Delta^2+o(\Delta^2)\\
\mathbb{C}_{1,2}(\tau) &= e^{-\tau(\alpha_1-\beta_{11})}a_{12,1}M\left[Z_1,1\right]M\left[Z_2,1\right]\Delta^2+o(\Delta^2)\\
\mathbb{C}_{2,1}(\tau) &= \left(e^{-\tau(\alpha_1-\beta_{11})}a_{21,1}+e^{-\tau(\alpha_2-\beta_{22})}a_{21,2}\right)M\left[Z_1,1\right]M\left[Z_2,1\right]\Delta^2+o(\Delta^2)\\
\mathbb{C}_{2,2}(\tau) &= \left(e^{-\tau(\alpha_1-\beta_{11})}a_{22,1}+e^{-\tau(\alpha_2-\beta_{22})}a_{22,2}\right)M\left[Z_2,1\right]^2\Delta^2+o(\Delta^2)
\end{aligned}$$

where

$$\begin{aligned}
a_{11,1} &= \tfrac{(2\alpha_1-\beta_{11})\beta_{11}\lambda_1}{2(\alpha_1-\beta_{11})}, \quad a_{12,1}=\tfrac{(2\alpha_1-\beta_{11})\beta_{11}\beta_{21}\lambda_1}{2(\alpha_1-\beta_{11})(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})}\\
a_{21,1} &= \tfrac{(2\alpha_1-\beta_{11})\beta_{11}\beta_{21}\lambda_1}{2(\alpha_1-\beta_{11})(-\alpha_1+\alpha_2+\beta_{11}-\beta_{22})}, \quad a_{21,2}=\tfrac{\beta_{21}\left(\alpha_1^2-(\alpha_2-\beta_{22})^2\right)\lambda_1}{(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})(\alpha_1-\alpha_2-\beta_{11}+\beta_{22})}\\
a_{22,1} &= \tfrac{(2\alpha_1-\beta_{11})\beta_{11}\lambda_1\beta_{21}^2}{2(\alpha_1-\beta_{11})(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})(-\alpha_1+\alpha_2+\beta_{11}-\beta_{22})}\\
a_{22,2} &= \tfrac{(2\alpha_2-\beta_{22})\beta_{22}\lambda_2}{2(\alpha_2-\beta_{22})}+\tfrac{\left(\alpha_1^2-(\alpha_2-\beta_{22})^2\right)\lambda_1\beta_{21}^2}{2(\alpha_2-\beta_{22})(\alpha_1+\alpha_2-\beta_{11}-\beta_{22})(\alpha_1-\alpha_2-\beta_{11}+\beta_{22})}.
\end{aligned}$$

As in the univariate case, the diffusive component of the model generates no autocorrelation. Also, the autocorrelation structure is asymmetric, reflecting the fact that the excitation from jumps in sector 1 to jumps in sector 2 is not the same as that from jumps in sector 2 to jumps in sector 1.

3.4 Interval-Based Moment Functions

GMM can be carried out using the instantaneous moments above, on short time intervals (daily). However, in order to improve accuracy and reduce the number of moment conditions to be used, we calculate in addition the unconditional interval-based moments ($i,j=1,2; 0\le s_1<s_2<s_3<s_4$)

$$\begin{cases}
\mathbb{E}\left[\int_{s_1}^{s_2}dX_{i,t}\right]\\
\mathbb{E}\left[\left(\int_{s_1}^{s_2}dX_{i,t}-\mathbb{E}\left[\int_{s_1}^{s_2}dX_{i,t}\right]\right)^2\right]\\
\mathbb{E}\left[\int_{s_1}^{s_2}dX_{i,t}\int_{s_1}^{s_2}dX_{j,u}-\mathbb{E}\left[\int_{s_1}^{s_2}dX_{i,t}\right]\mathbb{E}\left[\int_{s_1}^{s_2}dX_{j,u}\right]\right], \qquad i\neq j\\
\mathbb{E}\left[\int_{s_3}^{s_4}dX_{i,u}\int_{s_1}^{s_2}dX_{j,t}-\mathbb{E}\left[\int_{s_3}^{s_4}dX_{i,u}\right]\mathbb{E}\left[\int_{s_1}^{s_2}dX_{j,t}\right]\right]
\end{cases}$$

which can be applied on time intervals of arbitrary lengths. We are also able to compute these integrated moment conditions in closed-form.

The explicit formulae we derive for the moment conditions are contained in the Appendix. The lengthy closed-form expressions for the integrals $I_{i,j,N}^{\tau}$ and $I_{i,j,N}$ defined in the Appendix as well as for the derivatives of the moment conditions (needed to compute Ω; see (3.10) below) are not displayed in this paper to save space. These expressions are available in computer form from the authors upon request.

3.5 GMM Estimation

Given the expressions for the moment functions above, we proceed to estimate the parameters of the model using GMM. Let $Y_{n\Delta} := X_{n\Delta} - X_{(n-1)\Delta}$, $n = 1, \ldots, N$, $\Delta > 0$, with $N\Delta = T$, denote the log-returns on the interval $[0, T]$ and let $\theta \in \Theta$ denote our d-dimensional parameter vector. To estimate θ we consider a vector of M moment conditions $h(y, \Delta, \theta)$, $M \geq d$, continuously differentiable in θ. Let θ_0 denote the true value of θ and suppose that $\mathbb{E}\left[h(Y_{n\Delta}, \Delta, \theta_0)\right] = 0$. This is the key requirement for consistency of the GMM estimator. The closed-form expressions for the moments derived above ensure that it will be satisfied, since we will use for each component of the vector h the difference between the corresponding sample moment of the log-returns and its closed-form expression derived under the model.

We assume that the log-returns are stationary, which is the case for our model under the parameter restrictions discussed above. Let $g_T(y, \Delta, \theta)$ denote the sample average of $h(y_{n\Delta}, \Delta, \theta)$, that is $g_T(y, \Delta, \theta) := \frac{1}{N}\sum_{n=1}^{N} h(y_{n\Delta}, \Delta, \theta)$. Then the GMM estimator $\hat{\theta}_T$ is the value of $\theta \in \Theta$ that minimizes the quadratic form $g_T(y, \Delta, \theta)' W_T g_T(y, \Delta, \theta)$, where W_T is an $M \times M$ positive definite weight matrix assumed to converge in probability to a positive definite limit W. The system is exactly identified if $M = d$, in which case the choice of W_T becomes irrelevant.

The weight matrix W_T can be chosen optimally in the following way: Suppose that the process $\{h(Y_{n\Delta}, \Delta, \theta_0)\}_{n=-\infty}^{\infty}$ is strictly stationary with mean zero and v-th autocovariance matrix $\Gamma_v := \mathbb{E}\left[h(Y_{n\Delta}, \Delta, \theta_0)\ h(Y_{(n-v)\Delta}, \Delta, \theta_0)'\right]$. Assuming that these autocovariances are absolutely summable (i.e., that the sequence of partial absolute sums has a finite limit), let $S := \sum_{v=-\infty}^{\infty} \Gamma_v$. We note that S is the asymptotic variance of the sample average of $h(y_{n\Delta}, \Delta, \theta_0)$:

$$S = \lim_{N\to\infty} N\ \mathbb{E}\left[g_{T_N}(Y, \Delta, \theta_0) g_{T_N}(Y, \Delta, \theta_0)'\right]. \tag{3.7}$$

The optimal weight matrix is S^{-1}. A non-negative definite estimator of S is

$$\hat{S}_T = \hat{\Gamma}_{0,T} + \sum_{v=1}^{q} \left(1 - \frac{v}{q+1}\right)\left(\hat{\Gamma}_{v,T} + \hat{\Gamma}'_{v,T}\right), \tag{3.8}$$

with

$$\hat{\Gamma}_{v,T} = \frac{1}{N}\sum_{n=v+1}^{N} h(y_{n\Delta}, \Delta, \tilde{\theta})\ h(y_{(n-v)\Delta}, \Delta, \tilde{\theta})'. \tag{3.9}$$

Here $\tilde{\theta}$ is an initial consistent estimate of θ_0, which can be obtained by minimizing the quadratic form with $W_T = I$.

Under standard regularity conditions (see Hansen (1982)), $\sqrt{T_N}(\hat{\theta} - \theta_0)$ converges in law to

$N(0, \Omega)$ where

$$\Omega^{-1} := \Delta^{-1} D' W D (D' W S W D)^{-1} D' W D, \tag{3.10}$$

and D is the gradient of $E\left[h(Y_{n\Delta}, \Delta, \theta)\right]$ with respect to θ' evaluated at θ_0'. When W_T is chosen optimally, $W = S^{-1}$ and (3.10) reduces to $\Omega^{-1} = \Delta^{-1} D' S^{-1} D$.

3.6 Finite Sample Behavior and Practical Considerations

One advantage of the proposed closed-form estimation method is that it is numerically tractable, so that large numbers of Monte Carlo simulations can be conducted to determine the finite sample distribution of the estimators. In a setting designed to mimic the features of the empirical data analysis, we study the best choice of the array of autocovariances and cross-covariances to employ as moment functions, and the corresponding degree of accuracy of the estimation method for various sets of parameter values. We find that for daily data, the optimal array of moment conditions, striking a balance between accuracy and computational feasibility, is to use daily means and instantaneous (co)variances and to use daily, weekly or monthly autocovariances and cross-covariances depending on the degree of excitation. We find that using autocovariances and cross-covariances between periods of the same length rather than of different lengths produces the most stable results. Visual inspection of the autocorrelograms and cross-correlograms of the data is used to determine the number of autocovariances and cross-covariances we need to include in our GMM estimation. The Monte Carlo results are not included in the paper to save space but are available from the authors.

It is clear from the explicit expressions of the moments above that there will be difficulty in practice identifying the stochastic volatility parameters: there are simply too many latent processes between V_t and λ_t; while those are integrated out as part of the development of the unconditional moments, their corresponding parameters remain. So they are theoretically identified, but this identification can be tenuous in practice. This is an unavoidable consequence of the unobservability of those state variables. This will be especially so in the multivariate context where there are many potential correlation coefficients: between each asset's log-return and all the volatilities, among the different asset log-returns and among the different volatilities. To avoid this problem, we will restrict some of the parameters to be identical. In the bivariate case, the restrictions that $\alpha_1 = \alpha_2$ and $\lambda_{1,\infty} = \lambda_{2,\infty}$ facilitate the identification.

To disentangle diffusion from jumps we adopt a two-stage procedure: we first estimate the parameters $\{\mu_1, \mu_2, \theta_1, \theta_2, \rho\}$ from the continuous part of our model using truncated data. The truncation removes (most of) the jumps so that the truncated data can be viewed as being generated by the continuous part of our model. We fit the truncated data using moment conditions that are derived from the continuous part of our model, ignoring the discontinuous jump part. Due to the truncation, the volatility and correlation parameters are slightly biased downwards, which we correct for. Then in a second stage we treat the obtained parameter estimates for the continuous part of our model as fixed and given and identify the parameters of the Hawkes process $\{\gamma_1, \gamma_2, \lambda_\infty, \alpha, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$ using the full set of moment conditions. Using the starting values obtained in this way, we then estimate all the parameters of the Hawkes jump-diffusion model simultaneously using the full set of moments conditions (doing so removes the slight downward bias in the correlation and volatility parameters). We repeat this multiple times, using parameter estimates as starting values, and continuing until starting values coincide with corresponding parameter estimates and a minimal value function is obtained. Rather than using the standard identity matrix as the weight matrix in first stage GMM, we have found it helpful to put extra weights on the third and fourth moments to make them relatively comparable in magnitude to the other moments.

With these types of moment conditions and caveats, the estimates that we obtain from Monte Carlo simulated data are usually quite accurate. It behaves as is typical of GMM given the number of parameters of the model. Higher-dimensional versions of the model become more difficult to estimate, and the two-stage procedure helps in that it fixes the region of the parameter space over which to conduct the search. One approach to reduce the number of parameters consists in not having independent parameters for similar effects (for example, setting all the cross-excitation parameters to be identical). Another, which we have primarily adopted, consists in focusing only on fairly unrestricted but *bi*variate models. The Monte Carlo results show that the population parameters of the Hawkes jump-diffusion model can be identified with sufficient degree of precision from data generated by the Hawkes jump-diffusion model. We also analyzed the situation in which the data generating process is in fact a Poissonian jump-diffusion (with constant jump intensities) but is assumed to be a Hawkes jump-diffusion (with stochastic jump intensities). We find that our estimation methodology is able to distinguish between these two models, finding parameter estimates for $\alpha, \beta_{11}, \beta_{12}, \beta_{21}$ and β_{22} that are not significantly different from zero in this case.

4. Empirical Analysis

4.1 Data Sequencing

We are interested in capturing patterns of contagion between stock indices around the world. We use Morgan Stanley Capital International (MSCI) international equity index data. We will study six indices: US; Latin America (LA); UK; Developed countries Europe (EU); Japan (JA); Emerging markets Asia (ASEM). Daily data are available from January 1, 1980 (for US, UK, EU and JA), from January 29, 1988 (for LA) and from January 1, 1988 (for ASEM). Summary statistics are in Table 1. For all the regions, the excess kurtosis is substantially larger than that for a Gaussian distribution, as would be caused by jumps. Notice the difference in excess kurtosis between the US return series of panel A and the US return series of panel B. This difference is mainly due to the 1987 Black Monday, the returns of which are included in the longer sample period but not in the shorter sample period.

Autocorrelograms and cross-correlograms are in Figure 4. The model predicts that the empirical autocorrelation and cross-correlation is decreasing and convex, hence that the instantaneous correlation is the largest. The data reveals substantial correlations between US returns on day j and returns of other regions of the world on the following day $j+1$, except for Latin America where this phenomenon is less pronounced. It becomes apparent from the plots that autocorrelations and cross-correlations die out quickly, so that in five days time (or even fewer) most correlation has disappeared. The presence of these short-run correlations is what makes possible the identification of the self- and cross-excitation parameters.

In order to cope with the different time zones and hours of operation of the markets around the world some adjustments need to be made to the raw data to get the sequencing right given the time differences. To get a qualitative insight in the direction of jump transmissions, we start by sorting daily US returns to find the most extreme declines (over 3.0% in a single day) in the US stock market in the period January 1, 1980 to December 31, 2008. If the inter-arrival time of these jumps was less than 6 weeks, we grouped the returns as being one event, or a related sequence of events. We end up with 63 declines and 23 groups. We read the analysis in the press in the days following each event (statements such as "Tokyo opened lower *after* Wall Street closed down 3%" vs. "Wall Street opened lower *following* a rout in European markets") to confirm the sequencing: where and when the event started, and whether transmission (contagion) took place following one of these events. Generally speaking, we find that it is either (and primarily) a US news announcement that causes the US market to slump, and often such a decline transmits contagiously, or it is a non-US event, typically an adverse shock in emerging markets, and then other regions of the world decline once the US start declining. There are few exceptions to this general pattern of transmission.

For the US and Japan, and the US and Emerging Markets Asia, we calculate the instantaneous correlation by leading the Japanese return series by one day. To calculate the lagged cross-correlations we lead the Japanese return series by one day when computing $\mathrm{corr}(\Delta X_t^{\mathrm{US}}, \Delta X_{t+j}^{\mathrm{JA}})$, $j = 1, 2, \ldots$, and we do not lead the Japanese return series (i.e., use the original timing) when computing $\mathrm{corr}(\Delta X_t^{\mathrm{JA}}, \Delta X_{t+j}^{\mathrm{US}})$, $j = 1, 2, \ldots$. The same applies for the pair US and emerging markets Asia. This sequencing procedure is symmetric in the sense that both $\mathrm{corr}(\Delta X_t^{\mathrm{US}}, \Delta X_{t+1}^{\mathrm{JA}})$ and

$\text{corr}(\Delta X_t^{\text{JA}}, \Delta X_{t+1}^{\text{US}})$ correspond to a 36 hours lag.

For the US and UK, and the US and Developed Countries Europe, we calculate the instantaneous correlation by averaging the instantaneous and 1-day lagged UK returns. To calculate the lagged cross-correlations we average the j-day lagged and $j+1$-day lagged UK returns when computing $\text{corr}(\Delta X_t^{\text{US}}, \Delta X_{t+j}^{\text{UK}})$, $j = 1, 2, \ldots$, and we do not modify the UK series (i.e., use the original timing) when computing $\text{corr}(\Delta X_t^{\text{UK}}, \Delta X_{t+j}^{\text{US}})$, $j = 1, 2, \ldots$. The same applies for the pair US and developed countries Europe. We finally note that the pair US and Latin America does not require any modification because they operate in the same time zone.

4.2 Parameter Estimates

We use $4m + m(m-1)/2 + m^2 \times (\#\text{lags of autocovariances and cross-covariances})$ moment conditions to identify the parameters of our model, where in the bivariate case $m = 2$: the interval-based first moments, autocovariances and cross-covariances, and the third and fourth moments approximated at the leading order in Δ. Informed by Monte Carlo evidence, we restrict $\alpha_1 = \alpha_2 =: \alpha$ and that $\lambda_{1,\infty} = \lambda_{2,\infty} =: \lambda_\infty$, and we fix $p_1 = p_2 = 1$. We further use 3 lags of daily autocovariances and cross-covariances. This leaves us with 13 parameters to be identified by 21 moment conditions. GMM parameter estimates are in Table 2. One round of optimization (data import, GMM optimization, standard errors and test statistics) of the model takes a few minutes on a regular computer.

When computing the empirical third and fourth moments as well as the empirical lagged autocovariances and cross-covariances, the population counterparts of which are completely determined by the jump component of the model, we apply one-sided cutoffs to the data (of -2% for US, UK, EU and JA; -2.5% for ASEM; and -3% for LA), considering only the returns (likely jumps) that are smaller than the given cutoffs. The magnitudes of the cutoffs are dictated by inspection of the corresponding higher order moments and lagged autocovariances and cross-covariances, selecting values for which these moments are robust to changes in the cutoff magnitudes. Figure 5 plots the resulting empirical autocorrelations and cross-correlations. Visual inspection of the autocorrelograms and cross-correlograms is used to determine the number of autocovariances and cross-covariances we need to include (i.e., to determine the total length of the time period to be considered), and we initially estimate the model on this basis. Using the parameter values obtained in this way, we then compute the reconstructed autocovariances and cross-covariances from our model using the estimated parameter values to verify whether the autocovariances and cross-covariances beyond the number of lags considered are indeed approximately negligible.

We find large estimated values for β_{11} and β_{22}, measuring the degree of self-excitation. They provide clear evidence that both the US market as well as other markets strongly self-excite. These estimates come along with relatively moderate estimates for the volatility levels. It seems that, due to the integrated nature of the moment conditions, the jump part of our model partially captures what was traditionally (with continuous return dynamics) modeled as (instantaneous) volatility. Also, we typically find relatively large estimated values for β_{21}, measuring the degree of transmission from the US to other regions of the world. Large values for β_{21} imply that when the US jumps, there is a strong increase in the probability of a consecutive jump in another region of the world. From the empirical cross-correlation plots, the effect seems to be mainly driven by transmission on the same day or the day following the day of occurrence of a US jump.[8] Estimates of β_{12}, measuring the degree of the reverse transmission, are positive but relatively small. Statistically, the estimated values of the β_{12}'s are not significantly different from zero with daily data.

In addition to the parameter estimates, we also compute the model-implied autocovariances and cross-covariances that are obtained by substituting the parameter estimates into equations (B.7), (B.8) and (B.9), as well as the reconstructed third and fourth moments. Comparing the model-implied covariances and higher order moments with their empirical counterparts, we find that the model fits the data quite well.

4.3 Testing for the Presence of Various Forms of Contagion

If desired, it is possible to test for the presence of contagion in the context of our model. Contagion tests have been an important focus of the international finance literature (see Dungey, Fry, Gonzalez-Hermosillo, and Martin (2010) for a discussion of the existing tests in the literature). In the model, on a path-by-path basis, contagion takes the form of an increase in the likelihood of successive shocks in the affected countries beyond the initial shock and transmission. Whether this happens or not in our model is directly controlled by the presence or absence of the β parameters. More generally, testing for the presence of various forms of contagion is connected to testing the joint hypothesis that all the coefficients of mutual excitation $\beta_{i,j}$'s are 0. But it is also possible to test that any desired subset of the $\beta_{i,j}$'s are 0, leading to narrower definitions of contagion. For instance, we can further separate between testing for self- or time-series contagion (diagonal $\beta_{i,i} = 0$) as opposed to testing for cross-sectional contagion (off-diagonal $\beta_{i,j} = 0$, $i \neq j$). Or we could test for asymmetries in cross-sectional self-excitation, for instance testing that it flows in one direction only from a given market i (say, the US) to all the other markets but not in the reverse direction: this would correspond to a joint test of $\beta_{j,i} > 0$ and $\beta_{i,j} = 0$, for a fixed i and all $j \neq i$. Since the inference is based on standard

[8] Recall that, because of the sequencing procedure adopted (see Section 4.1), "the same day" for the pair US and JA (ASEM) means here date t for US and date $t+1$ for JA (ASEM). Similarly, "the day following the occurrence of a US jump" for the pair US and JA (ASEM) means here date t for US and date $t+2$ for JA (ASEM).

GMM given the relevant moment functions, GMM-based testing tools apply to test these or any other combination(s) of parameter restrictions.

We employ for this purpose the χ^2 statistic that follows from (3.10). The Hawkes jump-diffusion model nests the popular Poissonian jump-diffusion model, to which our model reduces when all $\beta_{i,j}$'s are 0 and therefore generates no contagion. In that case, we are testing the (contagion) model with Hawkes jumps and stochastic volatility, versus the nested (non-contagion) model consisting of Poissonian jumps and stochastic volatility. A Poissonian model with dependent jumps is non-nested in our model, but can still be tested using appropriate techniques for testing non-nested models along the lines of Vuong (1989). Specifically, we test the following null hypotheses: $\mathcal{H}_0^{\mathrm{I}} : \beta_{i,j} = 0,\ i, j = 1, 2$; $\mathcal{H}_0^{\mathrm{II}} : \beta_{i,i} = 0,\ i = 1, 2$; $\mathcal{H}_0^{\mathrm{III}} : \beta_{i,j} = 0,\ i, j = 1, 2,\ i \neq j$. We adopt a Wald chi-square test based on the GMM estimates. Test results are reported in Table 3. The null hypotheses are rejected in the vast majority of the cases, providing clear evidence for excitation (rejection of $\mathcal{H}_0^{\mathrm{I}}$), self-excitation (rejection of $\mathcal{H}_0^{\mathrm{II}}$) and cross-excitation (rejection of $\mathcal{H}_0^{\mathrm{III}}$). It means in particular that a Poisson jump-diffusion model is rejected when tested against the Hawkes jump-diffusion model. In addition to testing nested models with simpler jumps against our model, our estimation methodology also allows for testing the null hypothesis of simpler nested models with stochastic volatility but without jumps by considering $\mathcal{H}_0^{\mathrm{IV}} : \lambda_\infty = 0$. Table 2 shows that this null hypothesis is rejected in all cases.

Finally, we note that if the cross-excitation is asymmetric, which occurs when the off-diagonal parameters are not equal, say $\beta_{21} > \beta_{12}$, then the returns of asset 2 will appear to Granger-cause those of asset 1 if the two time series are subjected to a test of Granger causality. In the context of risk management, Granger causality has been proposed as a measure of the degree of connectivity between markets and their potential to generate systemic risk (see Billio, Getmansky, Lo, and Pelizzon (2010)).

4.4 Splitting Total Risk into Brownian, Poissonian Jump and Contagion Risks

As follows from the general theory of semimartingales, the second moments of log-returns given in Theorems 1 and 2 incorporate contributions to the quadratic variation of returns from all the components of the model. In economic terms, this means that we can split the total risk produced by an asset in terms of its Brownian (or continuous, or volatility) risk vs. jump risk. And because of the structure of the jumps, we can further split the jump risk into a Poissonian (or constant intensity) and a Hawkes (mutually exciting) component. The Hawkes component is the additional term in the variance equation $\mathbb{E}\left[\left(\int dX_t - \mathbb{E}\left[\int dX_t\right]\right)^2\right]$ that is $\beta-$dependent. Economically speaking, in light of the previous section, the percentage of the total risk that is due to the mutually exciting piece can be thought of as a contagion measure. Results of this split are reported in Table 4. We find that the mutual excitation component accounts for about 5% of the total risk as measured by the interval-based instantaneous autocovariance and about 10% of the total risk as measured by the interval-based instantaneous cross-covariance. Note that these are percentages of the total variability produced by an asset over its whole sample path in which jumps occur only rarely. It is moreover the component over and above the Poissonian component.

4.5 Robustness Checks: Open-to-Close and Close-to-Open Returns, and Sample Splits

To mitigate the need to carry out sequencing adjustments to the raw data, which depend on the particular time zone in which the markets are operating, we also carry out an analysis with open-to-close (= "daylight" return) and close-to-open returns (= "night" return). With a close-to-open return computed from the previous afternoon's closing price to the next morning's opening price, we have an instantaneous correlation to measure at the daily frequency. Daily returns can be computed either by summing the open-to-close return and the following close-to-open return or by summing the close-to-open return and the following open-to-close return. If the opening of the US market marks the beginning of a new day, the US daily returns will be computed by the former alternative while e.g., the Japanese daily returns will be computed by the latter alternative. If, however, the opening of the Japanese market marks the beginning of a new day, the US daily returns will be computed by the latter alternative while the Japanese daily returns will be computed by the former alternative. As we will see below these "natural" daily return definitions will introduce some asymmetry in the daily data.

We use open and close international equity index data from finance.yahoo.com. We study two indices: the S&P500 and Nikkei. Open-to-close and close-to-open data are available from January 4, 1984. We compare two situations, one where the opening of the US market marks the beginning of a new day, and one where the opening of the Japanese market marks the beginning of a new day. In the first, we find moderate correlation between Nikkei returns on day j and S&P500 returns on the following day $j+1$, while the correlation between S&P500 returns on day j and Nikkei returns on the following day $j+1$ is relatively small. Conversely, in the second situation we find substantial correlation between S&P500 returns on day j and Nikkei returns on the following day $j+1$, while the correlation between Nikkei returns on day j and S&P500 returns on the following day $j+1$ is approximately zero. This can be explained from the fact that in the first most transmissions from the US to Japan will already be reflected in the Japanese returns of the same day ("instantaneous"), hence a relatively small correlation between S&P500 returns on day j and Nikkei returns on the following day $j+1$. In the second, the transmissions from the US to Japan can only be reflected in the Japanese "night" returns of the same day or in the returns of the following days. It appears from the results that the Japanese night return does not significantly reflect the US cross-excitation, hence a substantial correlation between S&P500 returns on day j and Nikkei returns on the following day $j+1$. The converse applies to the transmission of shocks from Japan to the US.

Because the direction (causality) of transmissions on the same day cannot be identified with daily data (here composed from the intradaily open-to-close and close-to-open returns), GMM learns most about the direction of transmission from the covariances between day j and day $j+1$. The most natural way to proceed is therefore to let the opening of the Japanese market mark the beginning of a new day. The parameter estimates using these data are broadly consistent with the parameter estimates based on daily returns: there is clear evidence for self-excitation in the US market. The transmission from the US market to Japan is more pronounced than earlier, and the self-excitation in the Japanese market is still present, although a little less pronounced. Again, we find positive but small estimates, that are statistically not distinguishable from zero, for the reverse transmission from Japan to the US. We recall here that the sample period for the pair S&P500 and Nikkei is different from (shorter than) the sample period for the MSCI pair US and Japan considered earlier. Furthermore, while $\mathrm{corr}(\Delta X_t^{\mathrm{US}}, \Delta X_{t+1}^{\mathrm{JA}})$ and $\mathrm{corr}(\Delta X_t^{\mathrm{JA}}, \Delta X_{t+1}^{\mathrm{US}})$ correspond to a 36 hours lag, $\mathrm{corr}(\Delta X_t^{\mathrm{S\&P500}}, \Delta X_{t+1}^{\mathrm{Nikkei}})$ and $\mathrm{corr}(\Delta X_t^{\mathrm{Nikkei}}, \Delta X_{t+1}^{\mathrm{S\&P500}})$ correspond to a 24 hours lag. Again the tests provide clear evidence for excitation, self-excitation as well as cross-excitation (Table 3).

As a second robustness check, we have estimated the Hawkes jump-diffusion model over two subsamples of the full sample. Long time series are a necessary evil with any jump-based model since a sufficiently large number of jumps needs to have been observed in the sample to obtain meaningful estimation results. To assess the effect of sample changes on the parameter estimates, we have split the full sample into two subsamples: the first subsample covers January 1, 1980, to October 1, 2001, which corresponds to the first three quarters of the full sample ($[0, 3/4]$); the second subsample covers April 1, 1987, to December 31, 2008, which corresponds to the last three quarters of the full sample ($[1/4, 1]$). We focus attention on the pair (US,UK). The parameter estimates we find are again consistent: there is clear evidence for self-excitation in the US market and for cross-excitation from the US to the UK in both subsamples. We find that the expected number of jumps in the second subsample is estimated at larger values for both equity indices than the expected number of jumps in the first subsample. This is largely due to the fact that the second subsample covers the crisis episode in 2008, while this episode is not included in the first subsample.

4.6 Measuring Market Stress Using Filtered Values of the Jump Intensities

Stress in the marketplace is often measured using a volatility index such as the VIX from the Chicago Board Options Exchange (CBOE), which has often been labeled the "fear gauge." VIX levels above 80% at an annualized rate were recorded at the height of the financial crisis in the Fall of 2008, as shown in Figure 6. This is an unrealistic level from the perspective of a purely-diffusive model, at least if we take the implications of such a level over a few months' horizon, and is most likely indicative that jumps are feared. If the model does not allow for jumps, then the only way that high risk can be translated is through extreme volatility numbers. But what is measured by VIX is a form of total (continuous plus jump) variance. To the extent that the risk that is truly feared is that of jumps, then a measure that captures solely the jump risk would be advantageous.

Consider therefore the Hawkes jump intensities as a measure of market stress. By construction, higher values of $\lambda_{i,t}$ lead to higher probability of jumps in asset i, and should be reflected in higher derivative prices such as options. In a Hawkes model, these intensities are time-varying, and it is likely that they will reflect market conditions at the time. Just like volatilities, jump intensities are latent; unlike volatilities, no market instrument is currently traded which references jump intensity. There are no "jump intensity swaps," for instance. One solution to infer jump intensities is to filter them out of the observed time series of returns.

To illustrate, we focus on the US and UK markets. In order to filter jump intensities from the time series of asset returns, we insert the parameter estimates of Table 2 in (2.8), and estimate to be a jump (hence, $dN_{j,t} = 1$) each daily return in the corresponding time series that is below the one-sided cutoffs of -2%. Filtering jumps on the basis of large returns is a natural approach, see e.g., Lee and Mykland (2008). The resulting plots of the estimated time series of $(\lambda_{\text{US},t}, \lambda_{\text{UK},t})$ are in Figure 7. The plots show that the filtered $\lambda_{i,t}$'s are plausible indicators of market stress, increasing in particular in the sample around the crisis periods. Indeed, from the plots we observe that the two most noticeable periods of market stress are 2000-02 (Internet bubble burst, 9/11) and the fall of 2008 (credit crisis).

Once we identify the first jump in the time series (the market is stressed by this measure), the model predicts that further jumps are more likely to occur in the short run. We test whether this out-of-sample prediction dominates those from two alternative models: a model with stochastic volatility only, so all market stress is volatility-driven, and a model with Poissonian jumps, so the occurrence of a jump does not predict that more jumps are likely. We focus on the US. Specifically, we define a jump indicator taking the value one if the daily US index return is smaller than the one-sided cutoff of -2%, and zero otherwise. The predicted jump probability for the Hawkes jump-diffusion is $\hat{\lambda}_t \Delta$, with $\hat{\lambda}_t$ the filtered jump intensity and $\Delta = 1$ trading day. The predicted jump probability for the Poisson jump-diffusion is $\hat{\lambda}\Delta$, with $\hat{\lambda}$ the realized jump intensity. The predicted jump (or extreme decline) probability for the stochastic volatility model is approximated as the Gaussian probability of observing an equity index return smaller than -2%, given a realized volatility estimate (based on the full history or on a rolling window). The predictions are plotted in Figure 8. On the basis of the Mean Squared Prediction Error we find that the Hawkes jump-diffusion model clearly dominates the nested models with simpler jumps or without jumps.

5. Conclusions

We proposed an intuitively appealing reduced-form model for asset returns that is able to capture jump clustering in time and across assets, or financial contagion. The model is coined the *Hawkes jump-diffusion model*. Unlike other models based on Hawkes jump processes alone, our model incorporates the standard elements of drift and stochastic volatility in addition to these jumps, thereby being most naturally thought of as a generalization of the Poisson jump-diffusion model familiar to financial economists. We developed an estimation procedure for the model based on GMM. The estimation procedure relies entirely on closed-form moment functions, hence is amenable to a multivariate setting and is numerically tractable. Monte Carlo evidence shows the population parameters of the data generating process can be recovered with a sufficient degree of precision for practical applications. We implemented the estimation procedure on international equity data, studying the patterns of jump excitation among six world markets. Our empirical results indicate that both the US market as well as the other markets strongly self-excite. Furthermore, we find that US jumps tend to get reflected quickly in most other markets, while we find that statistical evidence for the reverse transmission is much less pronounced. The tests empirically reject a model based on Poissonian jumps against a model based on Hawkes jumps.

Appendix: Explicit Formulae for the Moment Conditions

Appendix A: Univariate Case

Let $m = 1$ and assume that

$$\begin{cases} dX_t = \mu dt + \sqrt{V_t} dW_t^X + Z_t dN_t \\ dV_t = \kappa(\theta - V_t)dt + \eta\sqrt{V_t} dW_t^V \\ d\lambda_t = \alpha\left(\lambda_\infty - \lambda_t\right)dt + \beta dN_t \end{cases} \tag{A.1}$$

with $\mathbb{E}\left[dW_t^X dW_t^V\right] =: \rho^V dt$. The vector of Brownian motions W, the jump size Z and the jump process N are assumed to be mutually independent. The corresponding integral equation for λ_t reads:

$$\lambda_t = \lambda_\infty + \int_{-\infty}^{t} \beta e^{-\alpha(t-s)} dN_s. \tag{A.2}$$

A.1 Moment Conditions: Univariate Self-Exciting Jumps

Write

$$\lambda := \mathbb{E}\left[\lambda_t\right], \qquad V_N(\tau) := \mathbb{E}\left[dN_{t+\tau} dN_t\right] / (dt)^2 - \lambda^2. \tag{A.3}$$

Then $\mathbb{E}\left[dX_t\right]/dt = \mu + \lambda\mathbb{E}\left[Z\right]$, and

$$\frac{\mathbb{E}[dX_{t+\tau} dX_t]}{(dt)^2} - \left(\frac{\mathbb{E}[dX_t]}{dt}\right)^2 = \left(\mathbb{E}\left[Z\right]\right)^2 V_N(\tau), \qquad \tau > 0; \tag{A.4}$$

with $\lambda = \lambda_\infty \alpha/(\alpha - \beta)$ and $V_N(\tau) = \frac{\beta\lambda(2\alpha-\beta)}{2(\alpha-\beta)} e^{-(\alpha-\beta)\tau}$ for $\tau > 0$. Furthermore,

$$\begin{aligned} \mathbb{E}\left[\left(dX_t - \mathbb{E}\left[dX_t\right]\right)^2\right]/dt &= \theta + \lambda\mathbb{E}\left[Z^2\right] \\ \mathbb{E}\left[\left(dX_t - \mathbb{E}\left[dX_t\right]\right)^3\right]/dt &= \lambda\mathbb{E}\left[Z^3\right] \\ \mathbb{E}\left[\left(dX_t - \mathbb{E}\left[dX_t\right]\right)^4\right]/dt &= \lambda\mathbb{E}\left[Z^4\right]. \end{aligned}$$

A.2 Interval-Based Moment Conditions: Univariate Self-Exciting Jumps

Let $0 \leq s_1 < s_2 < s_3 < s_4$ and let $\Delta_1 := s_2 - s_1$, $\Delta_2 := s_4 - s_3$ and $\tau := s_3 - s_1$. Write $\lambda := \mathbb{E}\left[\lambda_t\right]$, and

$$\begin{aligned} f_N(\tau, \Delta_1, \Delta_2) &:= \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dN_u \int_{s_1}^{s_2} dN_t\right]}{\Delta_1\Delta_2} - \lambda^2, \\ g_{N,Z}(\Delta_1) &:= \frac{\mathbb{E}\left[\left(\int_{s_1}^{s_2} Z_t dN_t\right)^2\right]}{\Delta_1} - \left(\mathbb{E}[Z]\right)^2 \lambda^2\Delta_1. \end{aligned}$$

Then $\mathbb{E}\left[\int_{s_1}^{s_2} dX_t\right] = \left(\mu + \lambda\mathbb{E}\left[Z\right]\right)\Delta_1$, and

$$\mathbb{E}\left[\int_{s_3}^{s_4} dX_u \int_{s_1}^{s_2} dX_t\right] - \mathbb{E}\left[\int_{s_3}^{s_4} dX_u\right]\mathbb{E}\left[\int_{s_1}^{s_2} dX_t\right] = \left(\mathbb{E}\left[Z\right]\right)^2 f_N(\tau, \Delta_1, \Delta_2)\Delta_1\Delta_2;$$

with

$$f_N(\tau, \Delta_1, \Delta_2) = \frac{\beta\lambda(2\alpha-\beta)}{2\Delta_1\Delta_2(\alpha-\beta)^3}\left(e^{-(\alpha-\beta)(\tau-\Delta_1)} - e^{-(\alpha-\beta)(\tau-\Delta_1+\Delta_2)} - e^{-(\alpha-\beta)\tau} + e^{-(\alpha-\beta)(\tau+\Delta_2)}\right).$$

Furthermore,

$$\mathbb{E}\left[\left(\int_{s_1}^{s_2} dX_t\right)^2\right] - \left(\mathbb{E}\left[\int_{s_1}^{s_2} dX_t\right]\right)^2 = \left(\theta + g_{N,Z}(\Delta_1)\right)\Delta_1;$$

with

$$g_{N,Z}(\Delta_1) = \lambda\mathbb{E}\left[Z^2\right] + \left(\mathbb{E}[Z]\right)^2 \frac{\beta\lambda(2\alpha-\beta)}{(\alpha-\beta)^2}\left(1 + \frac{1}{\Delta_1(\alpha-\beta)}\left(e^{-(\alpha-\beta)\Delta_1} - 1\right)\right).$$

Appendix B: Bivariate Case

Let $m = 2$ and assume that

$$\begin{cases} dX_{1,t} = \mu_1 dt + \sqrt{V_{1,t}} dW^X_{1,t} + Z_{1,t} dN_{1,t} \\ dX_{2,t} = \mu_2 dt + \sqrt{V_{2,t}} dW^X_{2,t} + Z_{2,t} dN_{2,t} \\ dV_{1,t} = \kappa(\theta_1 - V_{1,t}) dt + \eta_1 \sqrt{V_{1,t}} dW^V_t \\ dV_{2,t} = d\left(\frac{\theta_2}{\theta_1}\right) V_{1,t} \\ d\lambda_{1,t} = \alpha_1 \left(\lambda_{1,\infty} - \lambda_{1,t}\right) dt + \beta_{11} dN_{1,t} + \beta_{12} dN_{2,t} \\ d\lambda_{2,t} = \alpha_2 \left(\lambda_{2,\infty} - \lambda_{2,t}\right) dt + \beta_{21} dN_{1,t} + \beta_{22} dN_{2,t} \end{cases} \tag{B.1}$$

with $\mathbb{E}[dW^X_{1,t} dW^X_{2,t}] =: \rho dt$ and $\mathbb{E}[dW^X_{i,t} dW^V_t] =: \rho^V_i dt$, $i = 1, 2$. The vector of Brownian motions W, the vector of jump sizes Z and the vector of jump processes N are assumed to be mutually independent. The corresponding integral equation for $\lambda_{i,t}$ reads:

$$\lambda_{i,t} = \lambda_{\infty,i} + \int_{-\infty}^t \beta_{i,1} e^{-\alpha_i (t-s)} dN_{1,s} + \int_{-\infty}^t \beta_{i,2} e^{-\alpha_i(t-s)} dN_{2,s}, \qquad i = 1, 2. \tag{B.2}$$

B.1 Moment Conditions: Bivariate Mutually Exciting Jumps

Write

$$\lambda_i := \mathbb{E}\left[\lambda_{i,t}\right], \ i = 1, 2, \quad V_N(\tau) := \begin{pmatrix} \frac{\mathbb{E}[dN_{1,t+\tau} dN_{1,t}]}{(dt)^2} - \lambda_1^2 & \frac{\mathbb{E}[dN_{1,t+\tau} dN_{2,t}]}{(dt)^2} - \lambda_1 \lambda_2 \\ \frac{\mathbb{E}[dN_{2,t+\tau} dN_{1,t}]}{(dt)^2} - \lambda_1 \lambda_2 & \frac{\mathbb{E}[dN_{2,t+\tau} dN_{2,t}]}{(dt)^2} - \lambda_2^2 \end{pmatrix}. \tag{B.3}$$

Then $\mathbb{E}\left[dX_{i,t}\right]/dt = \mu_i + \lambda_i \mathbb{E}\left[Z_i\right]$, for $i = 1, 2$, and

$$\begin{aligned} &\begin{pmatrix} \frac{\mathbb{E}[dX_{1,t+\tau} dX_{1,t}]}{(dt)^2} - \left(\frac{\mathbb{E}[dX_{1,t}]}{dt}\right)^2 & \frac{\mathbb{E}[dX_{1,t+\tau} dX_{2,t}]}{(dt)^2} - \frac{\mathbb{E}[dX_{1,t}]}{dt} \frac{\mathbb{E}[dX_{2,t}]}{dt} \\ \frac{\mathbb{E}[dX_{2,t+\tau} dX_{1,t}]}{(dt)^2} - \frac{\mathbb{E}[dX_{1,t}]}{dt} \frac{\mathbb{E}[dX_{2,t}]}{dt} & \frac{\mathbb{E}[dX_{2,t+\tau} dX_{2,t}]}{(dt)^2} - \left(\frac{\mathbb{E}[dX_{2,t}]}{dt}\right)^2 \end{pmatrix} \\ &= \begin{pmatrix} \left(\mathbb{E}[Z_1]\right)^2 V_{1,1,N}(\tau) & \mathbb{E}[Z_1]\mathbb{E}[Z_2] V_{1,2,N}(\tau) \\ \mathbb{E}[Z_1]\mathbb{E}[Z_2] V_{2,1,N}(\tau) & \left(\mathbb{E}[Z_2]\right)^2 V_{2,2,N}(\tau) \end{pmatrix}, \qquad \tau > 0; \end{aligned} \tag{B.4}$$

with

$$\begin{aligned} \lambda_1 &= \frac{\lambda_{1,\infty}\alpha_1(\alpha_2 - \beta_{22}) + \lambda_{2,\infty}\alpha_2\beta_{12}}{(\alpha_1 - \beta_{11})(\alpha_2 - \beta_{22}) - \beta_{12}\beta_{21}}, \\ \lambda_2 &= \frac{\lambda_{2,\infty}\alpha_2(\alpha_1 - \beta_{11}) + \lambda_{1,\infty}\alpha_1\beta_{21}}{(\alpha_1 - \beta_{11})(\alpha_2 - \beta_{22}) - \beta_{12}\beta_{21}}; \end{aligned}$$

and

$$\begin{aligned} V_{1,1,N}(\tau) = \Big(&(1 + e^{r\,\tau}) \Big(2\,\beta_{11}\,\beta_{12}\,\beta_{21}\,\alpha_1\,\lambda_1 + 2\,\beta_{11}\,\alpha_1\,\left(\beta_{22} - \alpha_2\right)\left(-\beta_{22} + \alpha_1 + \alpha_2\right)\lambda_1 \\ &+ \beta_{12}^3\,\beta_{21}\,\lambda_2 - \beta_{12}^2\left(\beta_{21}^2\,\lambda_1 + \left(-\left(\beta_{22}\,\alpha_1\right) + \beta_{11}\,\left(\beta_{22} - \alpha_2\right) + \alpha_2\,\left(\alpha_1 + \alpha_2\right)\right)\lambda_2\right)\Big)\,r \\ &+ 2\,e^{\frac{r\,\tau}{2}}\,\Big(\beta_{11}^2\,\left(-\left(\beta_{12}\,\beta_{21}\right) + \left(\beta_{22} - \alpha_2\right)\left(\beta_{11} + \beta_{22} - 3\,\alpha_1 - \alpha_2\right)\right)\,\lambda_1\,r\,\cosh\left(\tfrac{r\,\tau}{2}\right) \\ &+ \left(\beta_{11} + \beta_{22} - \alpha_1 - \alpha_2\right)\Big(2\,\beta_{11}\,\left(\beta_{12}\,\beta_{21}\,\left(2\,\beta_{22} + \alpha_1 - 2\,\alpha_2\right)\right. \\ &+ \alpha_1\,\left(\beta_{22} - \alpha_2\right)\left(\beta_{22} + \alpha_1 - \alpha_2\right)\Big)\,\lambda_1 - \beta_{11}^2\,\left(\beta_{12}\,\beta_{21} + \left(\beta_{22} - \alpha_2\right)\left(\beta_{22} + 3\,\alpha_1 - \alpha_2\right)\right)\lambda_1 \\ &+ \beta_{11}^3\,\left(\beta_{22} - \alpha_2\right)\lambda_1 - \beta_{11}\,\beta_{12}^2\,\left(\beta_{22} - \alpha_2\right)\lambda_2 \\ &+ \beta_{12}\Big(-\left(\beta_{21}\,\left(3\,\beta_{12}\,\beta_{21} + 4\,\alpha_1\,\left(\beta_{22} - \alpha_2\right)\right)\lambda_1\right) \\ &+ \beta_{12}\,\left(\beta_{12}\,\beta_{21} + \beta_{22}\,\alpha_1 - \alpha_1\,\alpha_2 + \alpha_2^2\right)\lambda_2\Big)\Big)\,\sinh\left(\tfrac{r\,\tau}{2}\right)\Big)\Big) \\ /\Big(&4\,e^{\frac{(-\beta_{11} - \beta_{22} + \alpha_1 + \alpha_2 + r)\,\tau}{2}}\,\left(-\left(\beta_{12}\,\beta_{21}\right) + \left(\beta_{11} - \alpha_1\right)\left(\beta_{22} - \alpha_2\right)\right)\left(\beta_{11} + \beta_{22} - \alpha_1 - \alpha_2\right)r\Big); \end{aligned}$$

$$
\begin{aligned}
V_{1,2,N}(\tau) = \Big(& (-1 - e^{r\,\tau}) \Big(\beta_{21}\, (\beta_{12}\,\beta_{21}\,\alpha_1 - \beta_{11}\, (\beta_{12}\,\beta_{21} + 2\,\alpha_1\, (\beta_{22} - \alpha_2)))\ \lambda_1 \\
& +\beta_{12} \big((\beta_{22} - 2\,\alpha_1)\, (\beta_{12}\,\beta_{21} + \beta_{22}\,\alpha_1) \\
& - (\beta_{12}\,\beta_{21} + 2\, (\beta_{22} - \alpha_1)\, \alpha_1)\, \alpha_2 + 2\,\alpha_1\,\alpha_2^2 \\
& +\beta_{11}\, \left(2\,\beta_{12}\,\beta_{21} - \beta_{22}^2 + 2\,\beta_{22}\, (2\,\alpha_1 + \alpha_2) - 2\,\alpha_2\, (2\,\alpha_1 + \alpha_2)\right) \big)\, \lambda_2 \Big)\, r \\
& -2\, e^{\frac{r\,\tau}{2}}\, \Big(\beta_{11}^2\, (\beta_{22} - \alpha_2)\, (\beta_{21}\,\lambda_1 - 2\,\beta_{12}\,\lambda_2)\, r\, \cosh\left(\tfrac{r\,\tau}{2}\right) \\
& + \Big(\beta_{11}\,\beta_{21}\, \big(\beta_{12}\,\beta_{21}\, (3\,\beta_{22} + 2\,\alpha_1 - 3\,\alpha_2) \\
& +2\,\alpha_1\, (\beta_{22} - \alpha_2)\, (\beta_{22} + \alpha_1 - \alpha_2)\ \big)\, \lambda_1 - \beta_{12}\,\beta_{21}^2\, (2\,\beta_{12}\,\beta_{21} - \alpha_1\, (-3\,\beta_{22} + \alpha_1 + 3\,\alpha_2))\, \lambda_1 \\
& -\beta_{11}\,\beta_{12}\, \big(\beta_{22}^3 + \beta_{12}\,\beta_{21}\, (\beta_{22} + 4\,\alpha_1 - \alpha_2) - 6\,\alpha_1^2\,\alpha_2 + 2\,\alpha_2^3 + 2\,\beta_{22}\,\alpha_1\, (3\,\alpha_1 + 2\,\alpha_2) \\
& -\beta_{22}^2\, (2\,\alpha_1 + 3\,\alpha_2) \big)\, \lambda_2 + \beta_{12}\, \big(2\,\beta_{12}^2\,\beta_{21}^2 \\
& +\beta_{12}\,\beta_{21}\, \left(\beta_{22}^2 + \beta_{22}\, (\alpha_1 - 2\,\alpha_2) + (\alpha_1 - \alpha_2)\, (2\,\alpha_1 + \alpha_2)\right) \\
& +\alpha_1\, \left(\beta_{22}^3 + 2\,\beta_{22}\,\alpha_1\, (\alpha_1 + \alpha_2) - 2\, (\alpha_1 - \alpha_2)\, \alpha_2\, (\alpha_1 + \alpha_2) - \beta_{22}^2\, (\alpha_1 + 3\,\alpha_2)\right) \big)\, \lambda_2 \\
& +\beta_{11}^3\, (\beta_{22} - \alpha_2)\, (\beta_{21}\,\lambda_1 - 2\,\beta_{12}\,\lambda_2) \\
& +\beta_{11}^2\, \big(- (\beta_{21}\, (\beta_{12}\,\beta_{21} + (\beta_{22} - \alpha_2)\, (\beta_{22} + 3\,\alpha_1 - \alpha_2)\ )\, \lambda_1) \\
& +\beta_{12}\, \left(2\,\beta_{12}\,\beta_{21} - \beta_{22}^2 - 6\,\alpha_1\,\alpha_2 + 2\,\beta_{22}\, (3\,\alpha_1 + \alpha_2)\right)\, \lambda_2 \big) \Big)\, \sinh\left(\tfrac{r\,\tau}{2}\right)\Big)\Big) \\
/ \Big(& 4\, e^{\frac{(-\beta_{11} - \beta_{22} + \alpha_1 + \alpha_2 + r)\,\tau}{2}}\, \big(- (\beta_{12}\,\beta_{21}) + (\beta_{11} - \alpha_1)\, (\beta_{22} - \alpha_2) \big)\, (\beta_{11} + \beta_{22} - \alpha_1 - \alpha_2)\, r \Big);
\end{aligned}
$$

where

$$
r := \sqrt{\beta_{11}^2 + 4\beta_{12}\beta_{21} - 2\beta_{11}\left(\beta_{22} + \alpha_1 - \alpha_2\right) + \left(\beta_{22} + \alpha_1 - \alpha_2\right)^2},
$$

and $\tau > 0$. The expressions for $V_{2,2,N}(\tau)$ and $V_{2,1,N}(\tau)$ are obtained from these by interchanging suffices. Furthermore,

$$
\begin{aligned}
\frac{\mathbb{E}\left[(dX_{i,t})^2\right]}{dt} - \frac{(\mathbb{E}[dX_{i,t}])^2}{dt} &= \theta_i + \lambda_i \mathbb{E}\left[Z_i^2\right], \qquad i = 1, 2; \\
\frac{\mathbb{E}[dX_{1,t} dX_{2,t}]}{dt} - \frac{\mathbb{E}[dX_{1,t}]\mathbb{E}[dX_{2,t}]}{dt} &= \rho\sqrt{\theta_1\theta_2}.
\end{aligned}
\tag{B.5}
$$

B.2 Interval-Based Moment Conditions: Bivariate Mutually Exciting Jumps

Let $0 \leq s_1 < s_2 < s_3 < s_4$ and let $\Delta_1 := s_2 - s_1$, $\Delta_2 := s_4 - s_3$ and $\tau := s_3 - s_1$. Write

$$
\lambda_i := \mathbb{E}\left[\lambda_{i,t}\right], \ i = 1, 2, \quad V_N(\tau) := \begin{pmatrix} \frac{\mathbb{E}[dN_{1,t+\tau} dN_{1,t}]}{(dt)^2} - \lambda_1^2 & \frac{\mathbb{E}[dN_{1,t+\tau} dN_{2,t}]}{(dt)^2} - \lambda_1\lambda_2 \\ \frac{\mathbb{E}[dN_{2,t+\tau} dN_{1,t}]}{(dt)^2} - \lambda_1\lambda_2 & \frac{\mathbb{E}[dN_{2,t+\tau} dN_{2,t}]}{(dt)^2} - \lambda_2^2 \end{pmatrix}. \tag{B.6}
$$

Then $\mathbb{E}\left[\int_{s_1}^{s_2} dX_{i,t}\right] / \Delta_1 = \mu_i + \lambda_i \mathbb{E}\left[Z_i\right]$, for $i = 1, 2$, and

$$
\begin{aligned}
&\begin{pmatrix} \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dX_{1,u} \int_{s_1}^{s_2} dX_{1,t}\right]}{\Delta_1\Delta_2} - \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dX_{1,u}\right]}{\Delta_2} \frac{\mathbb{E}\left[\int_{s_1}^{s_2} dX_{1,t}\right]}{\Delta_1} & \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dX_{1,u} \int_{s_1}^{s_2} dX_{2,t}\right]}{\Delta_1\Delta_2} - \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dX_{1,u}\right]}{\Delta_2} \frac{\mathbb{E}\left[\int_{s_1}^{s_2} dX_{2,t}\right]}{\Delta_1} \\ \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dX_{2,u} \int_{s_1}^{s_2} dX_{1,t}\right]}{\Delta_1\Delta_2} - \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dX_{2,u}\right]}{\Delta_2} \frac{\mathbb{E}\left[\int_{s_1}^{s_2} dX_{1,t}\right]}{\Delta_1} & \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dX_{2,u} \int_{s_1}^{s_2} dX_{2,t}\right]}{\Delta_1\Delta_2} - \frac{\mathbb{E}\left[\int_{s_3}^{s_4} dX_{2,u}\right]}{\Delta_2} \frac{\mathbb{E}\left[\int_{s_1}^{s_2} dX_{2,t}\right]}{\Delta_1} \end{pmatrix} \\
&\quad = \begin{pmatrix} \frac{(\mathbb{E}[Z_1])^2 I_{1,1,N}^{\tau}(\tau,\Delta_1,\Delta_2)}{\Delta_1\Delta_2} & \frac{\mathbb{E}[Z_1]\mathbb{E}[Z_2] I_{1,2,N}^{\tau}(\tau,\Delta_1,\Delta_2)}{\Delta_1\Delta_2} \\ \frac{\mathbb{E}[Z_1]\mathbb{E}[Z_2] I_{2,1,N}^{\tau}(\tau,\Delta_1,\Delta_2)}{\Delta_1\Delta_2} & \frac{(\mathbb{E}[Z_2])^2 I_{2,2,N}^{\tau}(\tau,\Delta_1,\Delta_2)}{\Delta_1\Delta_2} \end{pmatrix};
\end{aligned}
\tag{B.7}
$$

with

$$
\begin{aligned}
I_{1,1,N}^{\tau}(\tau, \Delta_1, \Delta_2) &:= \textstyle\int_{u=s_3}^{s_4} \int_{t=s_1}^{s_2} V_{1,1,N}(u - t) du dt, \\
I_{1,2,N}^{\tau}(\tau, \Delta_1, \Delta_2) &:= \textstyle\int_{u=s_3}^{s_4} \int_{t=s_1}^{s_2} V_{1,2,N}(u - t) du dt.
\end{aligned}
$$

The expressions for $I_{2,2,N}^{\tau}$ and $I_{2,1,N}^{\tau}$ are obtained from these by replacing $V_{1,1,N}$ and $V_{1,2,N}$ by $V_{2,2,N}$ and $V_{2,1,N}$, respectively. Explicit expressions for $V_{i,j,N}(\tau)$ have been derived earlier; see Section B.1.

Furthermore,

$$\frac{\mathbb{E}\Big[\Big(\int_{s_1}^{s_2} dX_{i,t}\Big)^2\Big]}{\Delta_1} - \frac{\Big(\mathbb{E}\Big[\int_{s_1}^{s_2} dX_{i,t}\Big]\Big)^2}{\Delta_1} = \theta_i + \lambda_i \mathbb{E}\left[Z_i^2\right] + \frac{2(\mathbb{E}[Z_i])^2 I_{i,i,N}(\Delta_1)}{\Delta_1}, \qquad i = 1, 2; \tag{B.8}$$

$$\frac{\mathbb{E}\Big[\int_{s_1}^{s_2} dX_{1,t} \int_{s_1}^{s_2} dX_{2,u}\Big]}{\Delta_1} - \frac{\mathbb{E}\Big[\int_{s_1}^{s_2} dX_{1,t}\Big]\mathbb{E}\Big[\int_{s_1}^{s_2} dX_{2,u}\Big]}{\Delta_1} = \rho\sqrt{\theta_1\theta_2} + \frac{\mathbb{E}[Z_1]\mathbb{E}[Z_2]\big(I_{1,2,N}(\Delta_1)+I_{2,1,N}(\Delta_1)\big)}{\Delta_1}; \tag{B.9}$$

with

$$I_{i,i,N}(\Delta_1) := \int_{u=s_1}^{s_2} \int_{t=s_1}^{u} V_{i,i,N}(u-t)dudt, \qquad i = 1, 2,$$
$$I_{1,2,N}(\Delta_1) := \int_{u=s_1}^{s_2} \int_{t=s_1}^{u} V_{1,2,N}(u-t)dudt.$$

The expression for $I_{2,1,N}$ is obtained by replacing $V_{1,2,N}$ in the above expression by $V_{2,1,N}$.

References

Aït-Sahalia, Y., 2004, "Disentangling Diffusion from Jumps," *Journal of Financial Economics*, 74, 487–528.

Bates, D. S., 2000, "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 94, 181–238.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon, 2010, "Measuring systemic risk in the finance and insurance sectors," working paper, MIT.

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

Brillinger, D. R., 1988, "Some statistical methods for random process data from seismology and neurophysiology," *Annals of Statistics*, 16, 1–54.

Crane, R., and D. Sornette, 2008, "Robust dynamic classes revealed by measuring the response function of a social system," *Proceedings of the National Academy of Sciences*, 105, 15649–15653.

Dungey, M., R. Fry, B. Gonzalez-Hermosillo, and V. L. Martin, 2010, "Sampling Properties of Contagion Tests," working paper, University of Melbourne.

Dungey, M., and B. Gonzalez-Hermosillo, 2005, "Empirical modeling of contagion: A review of methodologies," *Quantitative Finance*, 5, 9–24.

Ehrmann, M., M. Fratzscher, and R. Rigobon, 2010, "Stocks, bonds, money markets and exchange rates: Measuring international financial transmission," *Journal of Applied Econometrics, forthcoming.*

Engle, R. F., 1982, "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation," *Econometrica*, 50, 987–1008.

Eraker, B., 2004, "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *The Journal of Finance*, 59, 1367Ű1404.

Errais, E., K. Giesecke, and L. R. Goldberg, 2010, "Affine point processes and portfolio credit risk," *SIAM Journal on Financial Mathematics, forthcoming.*

Hansen, L. P., 1982, "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

Hawkes, A. G., 1971a, "Point spectra of some mutually exciting point processes," *Journal of the Royal Statistical Society, Series B*, 33, 438–443.

——— , 1971b, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, 58, 83–90.

Hawkes, A. G., and D. Oakes, 1974, "A cluster representation of a self-exciting process," *Journal of Applied Probability*, 11, 493–503.

Heston, S., 1993, "A closed-form solution for options with stochastic volatility with applications to bonds and currency options," *Review of Financial Studies*, 6, 327–343.

Johannes, M., 2004, “The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models,” *The Journal of Finance*, 59, 227–260.

Lee, S., and P. A. Mykland, 2008, “Jumps in financial markets: A new nonparametric test and jump dynamics,” *Review of Financial Studies*, 21, 2535–2563.

Maheu, J. M., and T. H. McCurdy, 2004, “News arrival, jump dynamics and volatility components for individual stock returns,” *The Journal of Finance*, 59, 755–793.

Merton, R. C., 1976, “Option pricing when underlying stock returns are discontinuous,” *Journal of Financial Economics*, 3, 125–144.

Mohler, G. O., M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, 2011, “Self-Exciting Point Process Modeling of Crime,” *Journal of the American Statistical Association*, 106, 100–108.

Oakes, D., 1975, “The Markovian self-exciting process,” *Journal of Applied Probability*, 12, 69–77.

Ogata, Y., and H. Akaike, 1982, “On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes,” *Journal of the Royal Statistical Society, Series B*, 44, 102–107.

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

Reynaud-Bouret, P., and S. Schbath, 2010, “Adaptive estimation for Hawkes processes; Application to genome analysis,” *Annals of Statistics*, 38, 2781–2822.

Salmon, M., and W. W. Tham, 2007, “Time deformation and the yield curve,” working paper, Warwick Business School.

Vuong, Q. H., 1989, “Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses,” *Econometrica*, 57(2), 307–333.

Yu, J., 2004, “Empirical Characteristic Function Estimation and its Applications,” *Econometric Reviews*, 23, 93–123.

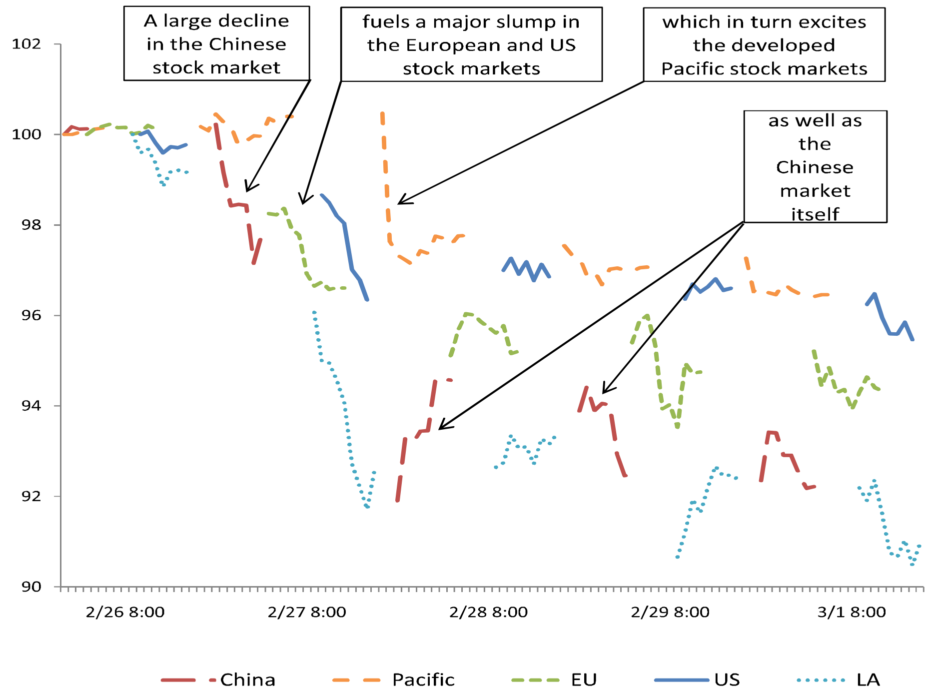

Figure 1
Mutual Excitation: Example I.

This figure plots the cascade of declines in international equity markets experienced between February 26, 2007 and March 1, 2007 in the US; Latin America (LA); Developed European countries (EU); China; and Developed countries in the Pacific. Data are hourly. The first observation of each of the price index series is normalized to 100 and the following observations are normalized by the same factor. Source: MSCI MXRT international equity indices on Bloomberg.

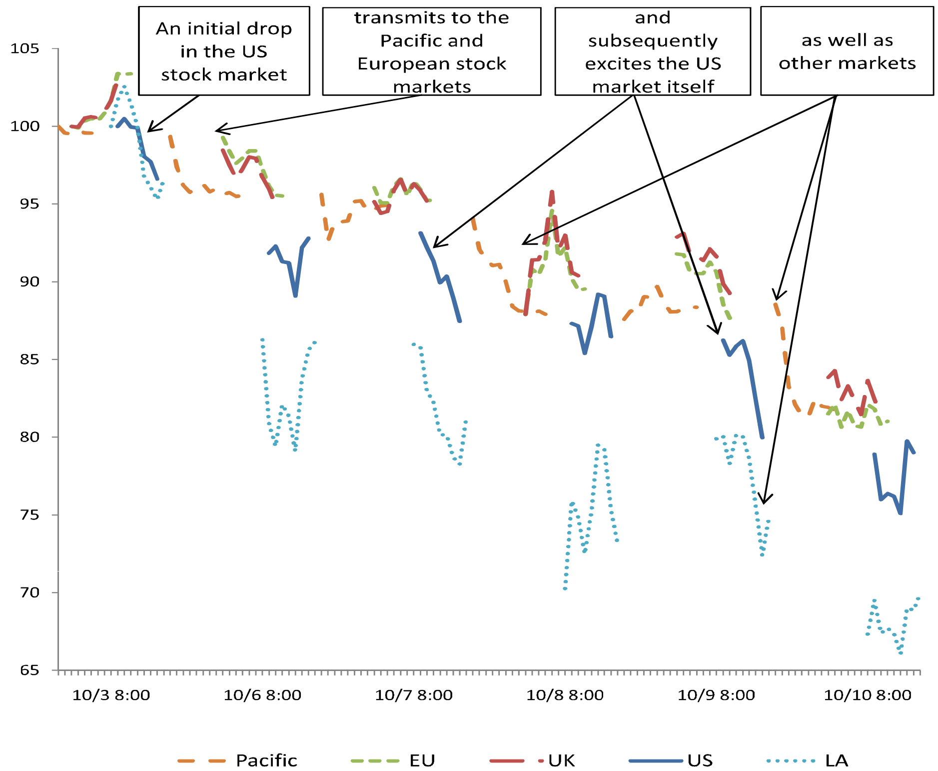

Figure 2
Mutual Excitation: Example II.

This figure plots the cascade of declines in international equity markets experienced between October 3, 2008 and October 10, 2008 in the US; Latin America (LA); UK; Developed European countries (EU); and Developed countries in the Pacific. Data are hourly. The first observation of each of the price index series is normalized to 100 and the following observations are normalized by the same factor. Source: MSCI MXRT international equity indices on Bloomberg.

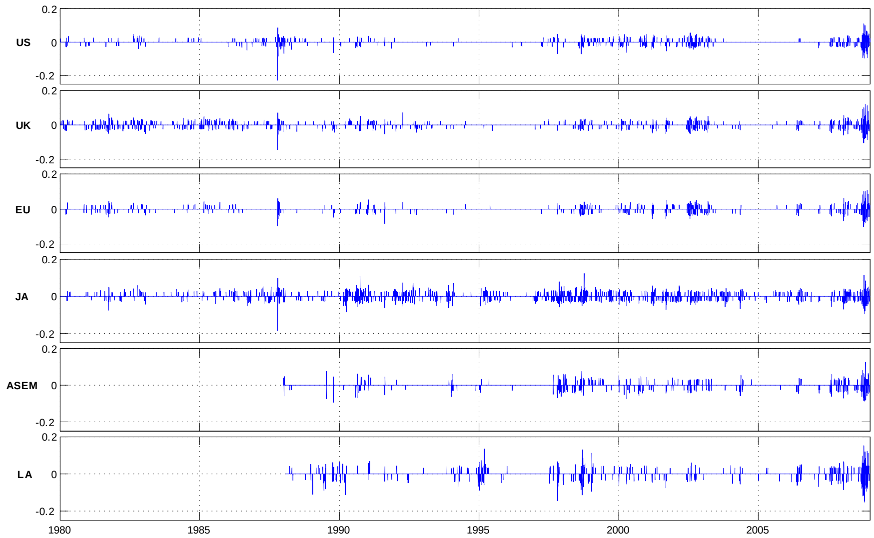

Figure 3
Large Stock Index Returns in the Six World Regions.

This figure plots the large log-returns of daily MSCI international equity index data. We consider six series: US; UK; Developed countries Europe (EU); Japan (JA); Emerging markets Asia (ASEM); and Latin America (LA). Sample period US, UK, EU and JA: January 1, 1980 to December 31, 2008; sample period ASEM: January 1, 1988 to December 31, 2008; sample period LA: January 29, 1988 to December 31, 2008. The large stock index returns plotted are obtained from the raw stock index returns by considering only those returns that are larger in absolute value than minus the cutoffs documented in Section 4.2. The purpose of the truncation is to illustrate the clustering of large returns. The descriptive statistics in Table 1 are based on the original (non-truncated) log-returns.

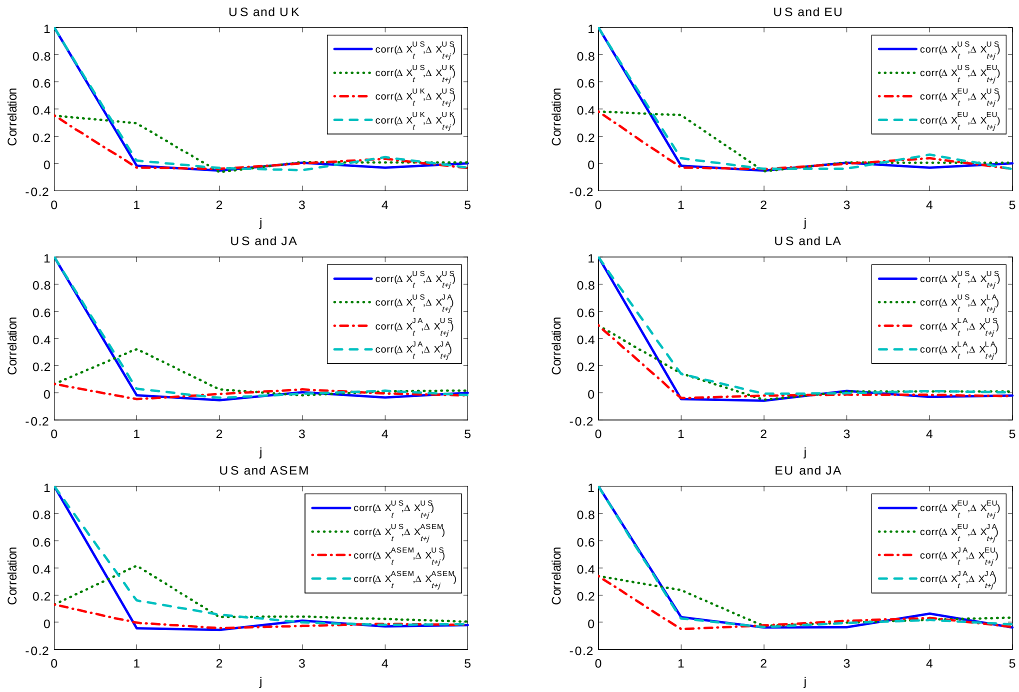

Figure 4
Autocorrelations and Cross-Correlations (Raw Data).

This figure plots autocorrelations and cross-correlations for the log-returns of the daily MSCI international equity data of Table 1. We consider six series: US; UK; Developed countries Europe (EU); Japan (JA); Emerging markets Asia (ASEM); and Latin America (LA). Sample period US, UK, EU and JA: January 1, 1980 to December 31, 2008; sample period ASEM: January 1, 1988 to December 31, 2008; sample period LA: January 29, 1988 to December 31, 2008. The unit of the index j is days.

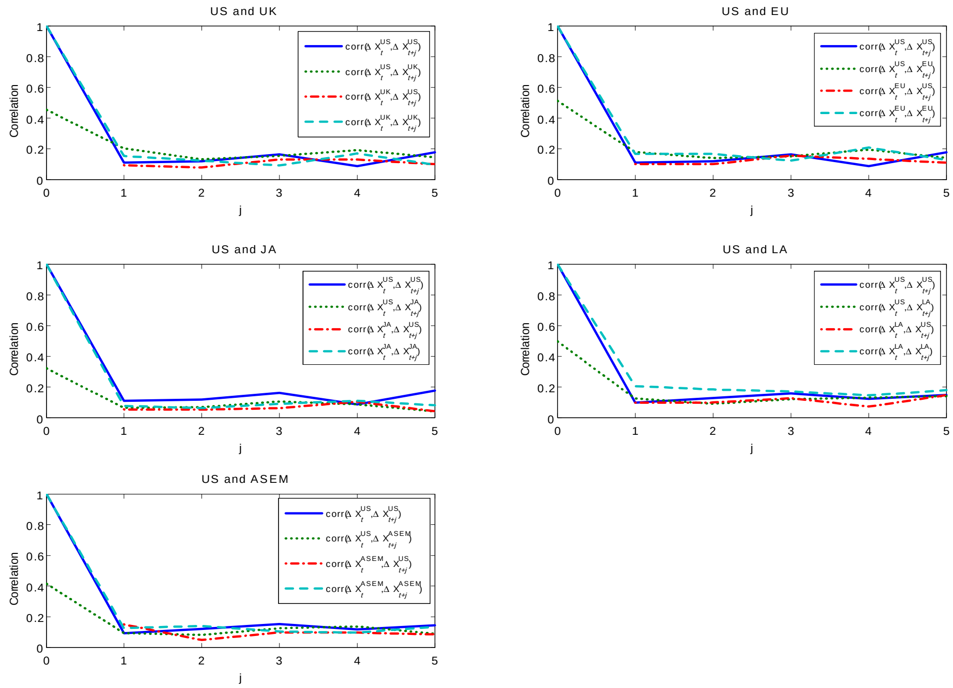

Figure 5
Autocorrelations and Cross-Correlations (Adjusted Data).

This figure plots autocorrelations and cross-correlations for the adjusted log-returns of the daily MSCI international equity data with cutoffs. The unit of the index j is number of days.

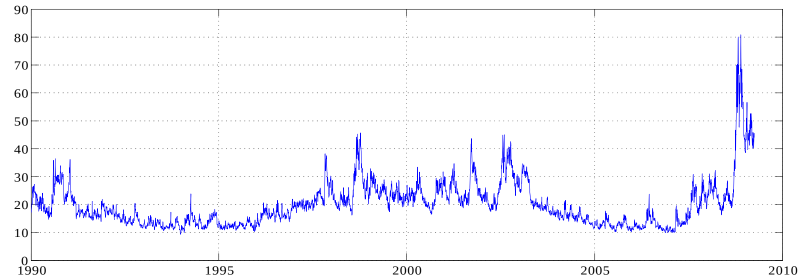

Figure 6
Volatility Index (VIX).

This figure plots the close index values of the VIX. Sample period: January 2, 1990 to March 31, 2009.

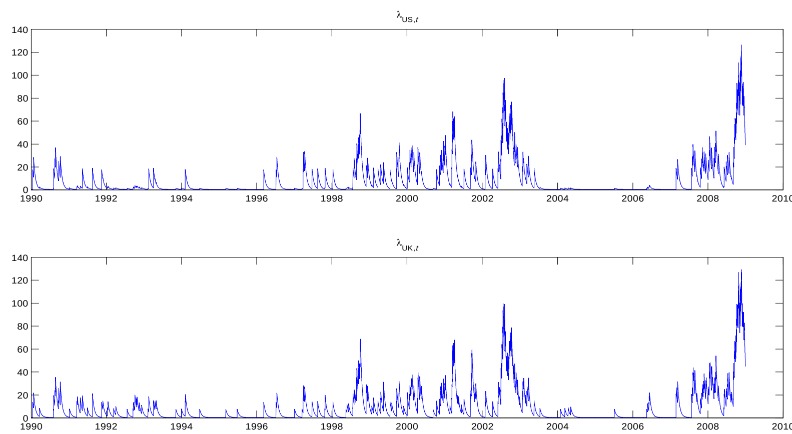

Figure 7
Estimated Time Series of the US and UK Jump Intensities.

This figure plots the estimated time series of $\lambda_{\mathrm{US},t}$ and $\lambda_{\mathrm{UK},t}$, filtered from the time series of asset returns. The adopted filtering procedure is described in Section 4.6. Sample period: January 2, 1990 to March 31, 2009.

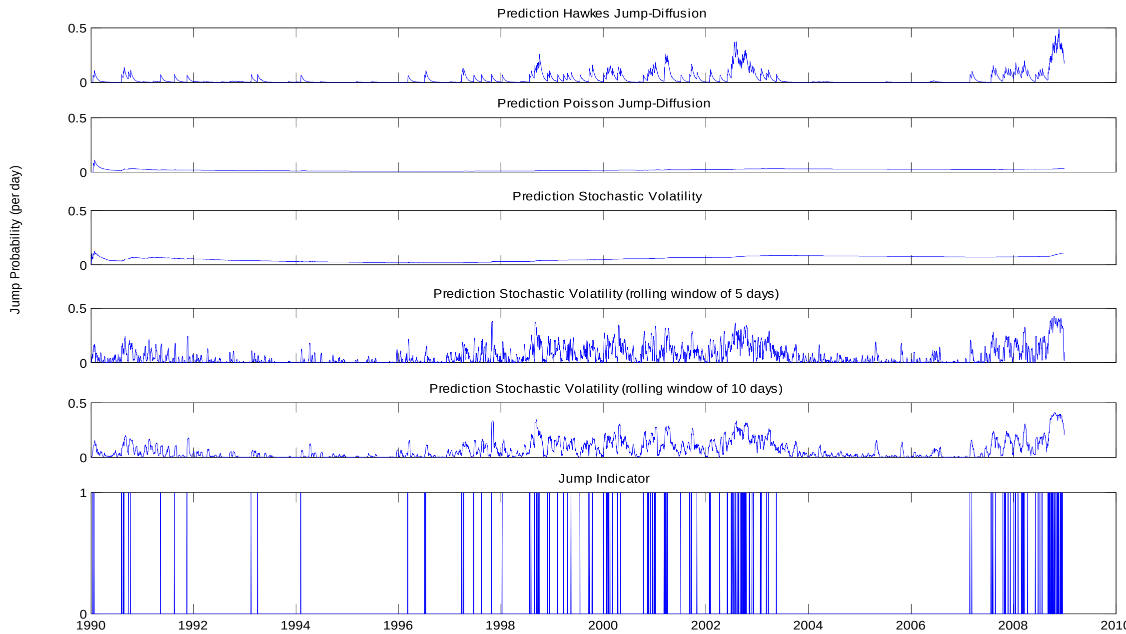

Figure 8
Out-of-Sample Prediction of Jump Probabilities.

This figure plots the out-of-sample prediction of the jump probability for a model with stochastic volatility only, a model with Poissonian jumps and a model with Hawkes jumps. The predicted jump probability for the Hawkes jump-diffusion is $\hat{\lambda}_t\Delta$, with $\hat{\lambda}_t$ the filtered jump intensity and $\Delta = 1$ trading day. The predicted jump probability for the Poisson jump-diffusion is $\hat{\lambda}\Delta$, with $\hat{\lambda}$ the realized jump intensity. The predicted "jump" (or extreme decline) probability for the stochastic volatility model is approximated as the Gaussian probability of observing an equity index return smaller than -2%, given a realized volatility estimate (based on the full history or on a rolling window). The jump indicator takes the value one if the daily US index return is smaller than the one-sided cutoff of -2%, and zero otherwise.

Panel A	US	UK	EU	JA
Mean	0.00028	0.00024	0.00027	0.00023
St. Deviation	0.011	0.012	0.011	0.014
Skewness	-1.39	-0.38	-0.37	-0.103
Excess Kurtosis	32.9	10.3	10.9	8.48

Panel B	US	LA	ASEM
Mean	0.00023	0.00053	0.00016
St. Deviation	0.011	0.018	0.013
Skewness	-0.28	-0.48	-0.31
Excess Kurtosis	10.7	10.6	7.3

Table 1
Descriptive Statistics: MSCI International Equity Indices.

This table reports descriptive statistics for the log-returns of daily MSCI international equity index data. Panel A: Sample period: January 1, 1980 to December 31, 2008. Panel B: Sample period US and Latin America: January 29, 1988 to December 31, 2008; sample period Emerging Markets Asia: January 1, 1988 to December 31, 2008.

1	US	US	US	US	US
2	UK	EU	JA	LA	ASEM
α	20.3**	21.8**	21.3***	24.2***	23.5***
	(9.3)	(9.2)	(6.5)	(4.9)	(0.8)
β_{11}	17.1**	18.6***	17.0**	17.8**	16.2***
	(6.7)	(6.2)	(7.9)	(7.6)	(4.9)
β_{12}	1.2	1.2	1.9	3.0	3.9
	(2.9)	(2.3)	(4.0)	(8.4)	(3.3)
β_{21}	13.1***	8.2**	4.0	5.6	8.6***
	(3.9)	(3.5)	(5.9)	(4.1)	(1.6)
β_{22}	7.1	13.6**	17.3***	18.5**	14.3***
	(4.9)	(6.8)	(5.0)	(7.6)	(1.8)
λ_∞	0.40***	0.31***	0.37**	1.80***	3.14***
	(0.081)	(0.066)	(0.159)	(0.252)	(0.156)
λ_1	4.63	3.88	4.81	19.83	29.54
λ_2	5.25	4.70	6.78	27.56	35.85
$\sqrt{\theta_1}$	0.146***	0.148***	0.146***	0.142***	0.138***
	(0.0132)	(0.0123)	(0.0153)	(0.0063)	(0.0049)
$\sqrt{\theta_2}$	0.175***	0.155***	0.199***	0.211***	0.164***
	(0.0078)	(0.0073)	(0.0059)	(0.0094)	(0.0052)
ρ	0.396***	0.456***	0.408***	0.757***	0.632***
	(0.0493)	(0.0401)	(0.0693)	(0.0748)	(0.0503)
μ_1	0.215***	0.199***	0.218***	0.370***	0.470***
	(0.0458)	(0.0412)	(0.0418)	(0.0395)	(0.0465)
μ_2	0.204***	0.188***	0.245***	0.774***	0.578***
	(0.0383)	(0.0346)	(0.0461)	(0.0751)	(0.0606)
$1/\gamma_1$	0.0307***	0.0326***	0.0304***	0.0156***	0.0138***
	(0.00726)	(0.00737)	(0.00669)	(0.00169)	(0.00138)
$1/\gamma_2$	0.0272***	0.0250***	0.0273***	0.0231***	0.0150***
	(0.00342)	(0.00306)	(0.00339)	(0.00189)	(0.00087)

Table 2
Parameter Estimates for the Bivariate Hawkes Jump-Diffusion Model.

This table reports the GMM parameter estimates for the 13 parameters of the bivariate Hawkes jump-diffusion model under exponential decay; standard errors are in parentheses. *, **, and *** indicate significance at the 95%, 97.5%, and 99.5% confidence levels, respectively. We consider six series: US; UK; Developed countries Europe (EU); Japan (JA); Emerging markets Asia (ASEM); and Latin America (LA). Sample period (US,UK), (US,EU) and (US,JA): January 1, 1980 to December 31, 2008; sample period (US,ASEM): January 1, 1988 to December 31, 2008; sample period (US,LA): January 29, 1988 to December 31, 2008.

	Null Hypothesis	US UK	US EU	US JA	US LA	US ASEM	S&P500 Nikkei
$\mathcal{H}_0^{\mathrm{I}}$	No Excitation	***	***	***	***	***	***
$\mathcal{H}_0^{\mathrm{II}}$	No Self-Excitation	***	***	***	***	***	**
$\mathcal{H}_0^{\mathrm{III}}$	No Cross-Excitation	***	***		**	***	***

Table 3
Contagion Tests Results.

This table reports the rejection significance of the Wald chi-square test statistics where the respective null hypotheses specify complete absence of excitation ($\beta_{i,j} = 0$, $i, j = 1, 2$), absence of self-excitation ($\beta_{i,i} = 0$, $i = 1, 2$), and absence of cross-excitation ($\beta_{i,j} = 0$, $i, j = 1, 2$, $i \neq j$). *, **, and *** indicate rejection at the 90%, 95%, and 99% confidence levels, respectively. We consider eight series: US; UK; Developed countries Europe (EU); Japan (JA); Emerging markets Asia (ASEM); Latin America (LA); S&P500; and Nikkei.

1 2	US UK	US EU	US JA	US LA	US ASEM	S&P500 Nikkei
Brownian auto 1	68.5%	68.9%	67.2%	65.3%	60.8%	59.1%
Poisson auto 1	26.7%	25.8%	28.2%	30.3%	35.0%	33.5%
Hawkes auto 1	4.8%	5.3%	4.6%	4.4%	4.3%	7.3%
Brownian cross 12	88.9%	90.2%	90.6%	92.0%	91.3%	84.0%
Poisson cross 12	0%	0%	0%	0%	0%	0%
Hawkes cross 12	11.1%	9.8%	9.4%	8.0%	8.7%	16.0%
Brownian auto 2	78.1%	77.4%	76.3%	56.9%	60.2%	79.1%
Poisson auto 2	18.8%	18.9%	19.6%	36.2%	35.1%	17.3%
Hawkes auto 2	3.1%	3.7%	4.1%	6.9%	4.7%	3.6%

Table 4
Splitting Total Risk into Brownian, Poissonian Jump and Hawkes Jump Components.

This table reports the split of the total risk produced by an asset in terms of its Brownian risk, its Poissonian component and its Hawkes component, on the basis of the interval-based instantaneous autocovariances and cross-covariances with the GMM parameter estimates of Table 2. The Hawkes component is the additional term in the autocovariance and cross-covariance equations that is $\beta-$dependent. We consider eight series: US; UK; Developed countries Europe (EU); Japan (JA); Emerging markets Asia (ASEM); Latin America (LA); S&P500; and Nikkei.