

# Modeling Financial Contagion

## Using Mutually Exciting Jump Processes\*

Yacine Aït-Sahalia	Julio Cacho-Diaz	Roger J. A. Laeven
Department of Economics	Department of Economics	Dept. of Quantitative Economics
Bendheim Center for Finance	Bendheim Center for Finance	University of Amsterdam
Princeton University and NBER	Princeton University	EURANDOM and CentER

This Version: August 16, 2014

### Abstract

We propose a model to capture the dynamics of asset returns, with periods of crises that are characterized by contagion. 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 observed features of the data. We develop and implement moment-based estimation and testing procedures for this model. 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 typically having more influence on the jump intensity of other markets than the reverse. We propose filtered values of the jump intensities as a measure of market stress and examine their out-of-sample forecasting abilities.

*Keywords:* Jumps; Contagion; Crisis; Hawkes process; Self- and mutually exciting processes.

JEL Codes: C58; G01; G15; C32.

---

\*We are very grateful for the comments of the Editor and an anonymous referee. We are also 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@uva.nl.

# 1. Introduction

Despite the predictions of standard continuous-time asset return models, financial crises seem to happen every decade or so. Indeed, 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.

Figures 1 and 2 illustrate two 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,<sup>1</sup> 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. 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.

Of course, jump processes can be employed to capture the large moves in asset markets that continuous, Brownian-based, models are unable to generate by themselves. But what makes a crisis worthy of that name 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. The interplay between the various jump terms across markets and over time is complex, and standard jump specifications appear to be unable to replicate those patterns.

In a typical jump model with arrival rates calibrated to historical data, jumps are inherently rare, occurring perhaps once a year on average. Hence observing patterns of multiple jumps in close succession over hours or days and across different markets, as in Figures 1 and 2, would be extremely unlikely under a standard jump model. More generally, over a longer time period, as in Figure 3 which shows all the large returns in five world regions' stock markets over the last twenty years, it is apparent that large moves tend to appear in clusters, both in time (horizontally)<sup>2</sup> and in space

---

<sup>1</sup>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.

<sup>2</sup>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 sixteen separate days. In the post-World War II era (until April 30, 2013, which is the end of our data sample), jumps of magnitude greater than 5% of either sign have only happened on twenty-three other days. And intra-day fluctuations were even more pronounced: during the same two months, the range of intra-day returns exceeded 10% during fourteen days. Jump clusters over time have been discussed in the financial econometrics literature: see Maheu and McCurdy (2004), Yu (2004) and Lee and Mykland (2008) in a univariate time series setting.

(vertically, across markets).<sup>3</sup>

The objective in this paper is to propose a model that is capable of reproducing both time and space propagation in a crisis, and develop appropriate estimation and testing methods for it. For this, we need to leave the widely applied class of Lévy jumps, such as the compound Poisson process that is the usual driving jump process 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 after Hawkes (1971b) (see also Hawkes (1971a), Hawkes and Oakes (1974) and Oakes (1975)). In a Hawkes process, a jump in one market raises the probability of future jumps both in the same market and elsewhere, thereby generating episodes that look just like those in 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).<sup>4</sup>

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.<sup>5</sup> 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

---

<sup>3</sup>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 occur simultaneously, especially when viewed at intra-daily 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. Clusters in space have been the focus of the contagion literature in international finance: see e.g., Ehrmann et al. (2010) and the review of Dungey and Gonzalez-Hermosillo (2005).

<sup>4</sup>Hawkes processes have been employed in a variety of applications, although always as pure jump processes. They 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 et al. (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 et al. (2011)).

<sup>5</sup>Other models can of course generate jump clustering. One prominent example is based on Cox processes, that is, a doubly stochastic Poisson process with stochastic intensity. However, such a model would not provide the feedback aspect of a Hawkes process and the intuitive causal interpretation of jumps causing more jumps; see Section 2.1 for further details. An econometric comparison of the different models that are able to generate jump clustering is outside the scope of this paper and is left to future research.

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 three contributions. First, we extend existing asset pricing models in two directions. Relative to existing applications of Hawkes processes, where they have been used as pure jump processes, our model adds a mutually exciting jump component to a continuous Brownian component with stochastic volatility, as well as a drift term. We refer to this model as a *Hawkes jump-diffusion* model by analogy with the Poissonian jump-diffusion model familiar to financial economists since Merton (1976). 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: to capture non-crisis periods, we also need a Brownian component, or alternatively, perhaps an infinite activity jump component, and the drift term of a semimartingale. And relative to the classical Merton (1976) model, we replace Poisson (or more generally Lévy) jumps with mutually exciting ones.

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, and we propose a new method to derive these estimating equations in closed-form. By contrast, the likelihood function of the model is not available in closed-form. And filtering methods based on high frequency data are not applicable in our context, due to the lack of intra-day data for many international stock indices.

Third, we estimate the model using international stock index returns for five 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-contagious) jumps is empirically rejected against the mutually exciting model based on Hawkes jumps, which allows for contagion. Finally, we propose to employ filtered jump intensities from the model as a measure of market stress, and examine their out-of-sample predictive power.

The rest of this paper is organized as follows: In Section 2, we present the model of asset return dynamics, and discuss some of its properties. In Section 3, we develop our estimation procedure and explain the calculation and selection of moment conditions and the identification strategy we employ.

Section 4 contains the empirical analysis. Conclusions are in Section 5.

## 2. Asset Return Dynamics

### 2.1. Jump Dynamics: Mutually Exciting Processes

Hawkes processes 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  $N_{i,t}$ ,  $i = 1, \dots, m$ , one for each of the  $m$  regions of the world.

Like 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

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

but instead of being constant the jump intensities follow the dynamics

$$\lambda_{i,t} = \lambda_{i,\infty} + \sum_{j=1}^m \int_{-\infty}^t g_{i,j}(t-s) dN_{j,s}, \quad i = 1, \dots, m. \quad (2)$$

In other words, each previous jump  $dN_{j,s}$ ,  $s \in (-\infty, t)$ ,  $j = 1, \dots, m$ , raises the jump intensities  $\lambda_{i,t}$ . Taken jointly,  $(N, \lambda)$  can be made into a Markov process by specializing the  $g(\cdot)$  function, as we do below. 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  $\sigma$ -algebra in (1) is  $\{N_{i,s}(s \leq t) : \lambda_{i,s}(-\infty < s < \infty)\}$ , i.e., the path of  $\lambda_{i,t}$  is not affected by the path of  $N_{i,t}$ .

Indeed, 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. Among state-dependent jump intensity models, what is gained by *self*- and *cross-excitation*, constituting mutual excitation in time and

space, relative to alternatives is: (i) Feedback – the occurrence of the jumps themselves is what is causing the higher rate of occurrence of future jumps, matching the empirical notion that market sentiment is moved in part by observed market moves, especially large ones. Price jumps themselves are heavily reported and often represent information that can generate further jumps. Once started, the cascade of jumps can be self-feeding (due to credit and liquidity shocks, margin calls, and so on) and contagious. Mutual excitation provides a natural model for capturing contagion induced by how financial markets are functioning themselves. (ii) Asymmetry – under mutual excitation, it is possible for only one of the countries' jumps to cause a pronounced increase in another country's jump intensity, capturing the asymmetry observed in the data. By contrast, suppose that the jump components of two countries are driven by exogenous state-dependent jump intensities that are linearly correlated through a common factor. Then, a jump in the common factor always raises the conditional probabilities of jumps in both countries. (iii) Parsimony – no additional exogenous state variables are needed to drive the jump intensities, unlike alternative models based e.g., on doubly stochastic Poisson processes. (iv) Realism – the model generates the jump clustering that is observed in the data; unless exogenous state variables are extremely autocorrelated and jump themselves, it is not possible using exogenous state variables in intensities to have jump intensities go from, say, 1 jump every year on average to 1 jump per week and stay there for a few days then mean-revert, matching the empirical observations on jump clusters. If they did, they would look very much like the jumps themselves, raising the question of why not using the price jumps directly, in view of parsimony.

We assume that in (2) the constant parameters  $\lambda_{i,\infty} \geq 0$  for all  $i = 1, \dots, m$ , that the real-valued functions  $g_{i,j}(u) \geq 0$  for all  $u \geq 0$  and for all  $i, j = 1, \dots, m$ . Notice that  $\lambda_{i,\infty}, g_{i,j}$  being non-negative for all  $i, j = 1, \dots, m$  ensures the non-negativity of the intensity processes with probability one. Let  $\Lambda_\infty$  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  $\lambda_i := \mathbb{E}[\lambda_{i,t}]$  denote the unconditional expected jump intensity, and  $\Lambda$  the  $m \times 1$  vector with components  $\lambda_i$ . Since from (1),  $\mathbb{E}[dN_{i,s}] = \lambda_i ds$ , we see that

$$\lambda_i = \lambda_{i,\infty} + \sum_{j=1}^m \lambda_j \int_{-\infty}^t g_{i,j}(t-s) ds = \lambda_{i,\infty} + \sum_{j=1}^m \left( \int_0^\infty g_{i,j}(u) du \right) \lambda_j, \quad (3)$$

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

## 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

$$dX_{i,t} = \mu_i dt + \sigma_i dW_{i,t} + Z_{i,t} dN_{i,t}, \quad i = 1, \dots, m, \quad (4)$$

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

$W_t := [W_{1,t}, \dots, W_{m,t}]'$  is an  $m$ -dimensional vector of standard Brownian motions with constant correlation coefficients  $\rho_{i,j}$ ,  $i, j = 1, \dots, m$ ,  $Z_t := [Z_{1,t}, \dots, Z_{m,t}]'$  is the vector of jump sizes, cross-sectionally and serially independently distributed with laws  $F_{Z_i}$  supported on  $(-\infty, \infty)$ , and  $N_t := [N_{1,t}, \dots, N_{m,t}]'$  is the vector of Hawkes processes just described.  $Z_i$  may be replaced by  $\log(1 + \tilde{Z}_i)$  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  $\log(1 + \tilde{Z}_i)$ . 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. Note that the model is based on jumps with finite activity, in contrast to the recent literature on high frequency financial statistics, which has been arguing in favor of infinite activity jump processes. This is because we are primarily interested in modeling (bi)daily returns, and employ the jumps to capture crises, which are (fortunately) rare events, rather than a large number of small jumps.

In the base model (4), the quantities  $\mu_i$  and  $\sigma_i$  are constant parameters. In this case, we assume that  $\Sigma$ , 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 (4), we allow for stochastic volatility:

$$dX_{i,t} = \mu_i dt + \sqrt{V_{i,t}} dW_{i,t}^X + Z_{i,t} dN_{i,t}, \quad (5)$$

where the instantaneous variance follows the Heston (1993) model

$$dV_{i,t} = \kappa_i(\theta_i - V_{i,t})dt + \eta_i \sqrt{V_{i,t}} dW_{i,t}^V, \quad (6)$$

with  $\kappa_i$ ,  $\theta_i$ ,  $\eta_i$  constant parameters.<sup>6</sup> Model (4) corresponds to the special case of (5)-(6) where the

---

<sup>6</sup>The Heston model for the stochastic volatility part of the model can be replaced by any other stochastic volatility model (with the moments below suitably adjusted). The main practical constraint in making that part of the model richer is one of identification of the additional parameters, such as those involved in stochastic correlations for instance, given the limitations of the data. In practice, not every part of the model can be made more general at the same time.

initial value is  $V_{i,0} = \theta_i$  and  $\eta_i = 0$ . In (6),  $V_{i,t}$  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 (5) as well as between the individual Brownian motions driving equations (5) and (6), thereby capturing a potential leverage effect. But in the presence of systematic jumps, the Brownian correlation, even if it increases in crisis episodes, will not play a major role: in times of crisis, jumps swamp everything else.<sup>7</sup> It is also possible for other channels to be involved in the contagion, such as the volatilities (as in Corradi et al. (2012)) but for reasons of tractability and practical identification of the parameters we restrict the excitation phenomenon to jumps in asset prices.

*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. This could be generated instead by *stochastic correlations* in a multivariate stochastic volatility model. Isolating a change in the structure of the Brownian variance-covariance matrix  $\Sigma$  becomes difficult in the context of a jump-diffusion model, 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. Indeed, one could alternatively assume that there is a common volatility factor that drives all international indices which would make volatility rise in all markets at the same time, and all normal shocks get amplified, leading to a contagion effect, although not of a magnitude of extreme returns and rate of occurrence compatible with what is actually observed in the data.

Given that the magnitude of such pure volatility moves is not sufficient to capture the observed index returns, adding Poisson jumps in variance as suggested in the extant literature would lead to spikes in volatility. But the financial econometrics literature also suggests that volatility spikes alone, even with jumps in volatility, are not sufficient to capture all large observed returns, and that

---

<sup>7</sup>More specifically, the claim is that in the presence of jumps that are correlated between the price and volatility processes, jumps are likely to drive the realized correlation. Suppose that the jumps are negatively correlated. To make an example, suppose that the price and volatility changes are respectively  $\{1, -1, 1, -1, 10\}$  and  $\{-1, 1, -1, 1, -5\}$  with the last observation, much larger than the other ones, being a jump. The contribution to the covariation of  $10 \times 5$  due to a jump is much larger than the products  $1 \times 1$ . In a normal period, there are few jumps and the contributions of the Brownian and jump components are, loosely, of the same order of magnitude. But in a crisis, there are more jumps and if there is a leverage effect in the jumps (meaning a negative correlation between jumps in price and jumps in volatility) then there will be relatively  $10 \times 5$  products in the covariance sum, which will swamp the  $1 \times 1$  products arising from the leverage effect in the Brownian components of the price and volatility.



jumps in asset prices are necessary: Any price process with only a stochastic volatility, including one that jumps, will always generate continuous price paths. However, the existence of large and infrequent jumps – the type included in this paper – is a rather robust finding, documented using a variety of different methods, ranging from parametric to nonparametric ones, and within each class of methods documented with many different estimators and testing procedures including: properties of the tail of the distribution of returns (Aït-Sahalia (2002)), evidence in higher moments (Aït-Sahalia (2004)), detection tests for large returns (Lee and Mykland (2008) and Lee and Hannig (2010)), bipower variations (Barndorff-Nielsen and Shephard (2006) and Huang and Tauchen (2005)), power variations (Aït-Sahalia and Jacod (2009) and Fan and Fan (2011)), truncation methods (Corsi et al. (2010)), evidence from options returns regarding jumps in the underlying asset price (Carr and Wu (2003)), evidence from dynamic replication deviations (Jiang and Oomen (2008)), etc. There is some debate in the literature regarding the reality of small and frequent jumps but we are not aware of any evidence in the literature claiming that there are no large and infrequent price jumps in the data. The fact that there are observed jumps in prices implies that the model must contain more than just stochastic volatility, no matter how sophisticated the model for stochastic volatility.

While having all these features present together would be quite useful, the model must be restricted if successful identification of the parameters is to be achieved in practice, given current technology and data limitations. Our interest here, and the novelty in this paper, consist in studying the role of mutual excitation in price jumps, making the necessary simplification of the other parts of the model where required. A comparison of models with different channels of transmission, such as volatility vs. jumps, is left to future research.

### 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

$$g_{i,j}(t-s) = \beta_{i,j}e^{-\alpha_i(t-s)}, \quad s < t, \quad i, j = 1, \dots, m, \quad (7)$$

with  $\alpha_i > 0$ ,  $\beta_{i,j} \geq 0$  for all  $i, j = 1, \dots, m$ , which results in  $(N, \lambda)$  being a Markov process. Under this model, 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  $\alpha_i$ . 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  $\beta_{i,j}$ ,  $i \neq j$ ).

The specification (7) plays an important role in allowing us to summarize the path-dependency of the model introduced by the history of the jumps into additional time- $t$  state variables, namely the jump intensities. With this model, the  $\Gamma$  matrix is given by  $\Gamma = [\beta_{i,j}/\alpha_i]_{i,j=1,\dots,m}$ . Under exponential decay (7), each jump intensity has the mean-reverting dynamics

$$d\lambda_{i,t} = \alpha_i (\lambda_{i,\infty} - \lambda_{i,t}) dt + \sum_{j=1}^m \beta_{i,j} dN_{j,t}. \quad (8)$$

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 (in time). 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 (in space).

#### 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  $M[Z, k] := \mathbb{E}[Z_t^k]$ . 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

$$F_{Z_i}(x) = \begin{cases} p_i e^{-\gamma_{i,1}(-x)}, & -\infty < x \leq 0; \\ p_i + (1 - p_i)(1 - e^{-\gamma_{i,2}x}), & 0 < x < \infty; \end{cases} \quad (9)$$

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

$$\mathbb{E}[Z_i^k] = (-1)^k \frac{k! p_i}{\gamma_{i,1}^k} + \frac{k! (1 - p_i)}{\gamma_{i,2}^k}, \quad k = 1, 2, \dots \quad (10)$$

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 now develop a GMM-based estimation procedure for the Hawkes jump-diffusion model. This model does not have a closed-form likelihood<sup>8</sup>, but we are able to derive closed-form expressions for the key moments of the log-returns. Having explicit moment functions 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. But deriving the explicit expressions of the moment functions is not trivial. Indeed, neither the point processes  $N_{i,t}$  and the intensity processes  $\lambda_{i,t}$ , nor the stochastic volatilities  $V_{i,t}$ ,  $i = 1, \dots, m$ , are directly observable. Instead, what is observed are the log-returns  $X_{i,t}$ .

The idea is to 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.

As is often the case, the selection of the moment conditions (or estimating equations) requires careful analysis. The moment conditions that we employ are:

$$\begin{cases} \mathbb{E} [\Delta X_{i,t}] \\ \mathbb{E} [(\Delta X_{i,t} - \mathbb{E} [\Delta X_{i,t}])^r], & r = 2, \dots, 4 \\ \mathbb{E} [\Delta X_{i,t} \Delta X_{j,t} - \mathbb{E} [\Delta X_{i,t}] \mathbb{E} [\Delta X_{j,t}]], & i \neq j \\ \mathbb{E} [(\Delta X_{i,t+\tau})^r (\Delta X_{j,t})^r - \mathbb{E} [(\Delta X_{i,t})^r] \mathbb{E} [(\Delta X_{j,t})^r]], & r = 1, 2, \quad \tau > 0. \end{cases} \quad (11)$$

The reason for using these moment functions is two-fold. First, they are “natural” in the sense of being easily interpretable and relevant in the context of financial applications: variance, skewness, kurtosis, autocovariances of returns and autocovariances of squared returns. Second, and more

---

<sup>8</sup>The process is Markovian when the state vector includes all the state variables, including the observable ones (log-returns) and the unobservable ones (volatilities and intensities). The log-likelihood function is then the sum of the logs of the transition densities of the full state vector evaluated at consecutive data points. The transition density solves a partial differential-difference equation (PDDE). These equations are more difficult to solve than partial differential equations (PDE), seldom have closed-form solutions, and the one for the model we consider doesn’t, to the best of our knowledge.

importantly, each moment function plays a specific role in identifying a specific part of the model. This is particularly crucial when estimating higher-dimensional models, since this feature means that we can decompose a single problem with many parameters into smaller-dimensional ones, thereby facilitating the estimation in practice.

### 3.1. Explicit Expressions for the Moments: Markov Infinitesimal Generator

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 (4) or (5)-(6) driven by the processes  $(W, N, \lambda)$ . Remarkably, this results in closed-form expressions for the moment functions of interest, which greatly simplifies the practical estimation.

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  $\delta$ . 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

$$\begin{aligned} \mathcal{A} \cdot g = & \frac{\partial g}{\partial \delta} + \sum_{i=1}^m \mu_i^X \frac{\partial g}{\partial y_{1,i}} + \sum_{i=1}^m \mu_i^V \frac{\partial g}{\partial v_{1,i}} + \sum_{i=1}^m \mu_i^\lambda \frac{\partial g}{\partial l_{1,i}} \\ & + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \sigma_{ij}^{YY} \frac{\partial^2 g}{\partial y_{1,i} \partial y_{1,j}} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \sigma_{ij}^{YV} \frac{\partial^2 g}{\partial y_{1,i} \partial v_{1,j}} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \sigma_{ij}^{VV} \frac{\partial^2 g}{\partial v_{1,i} \partial v_{1,j}} \\ & + \sum_{i=1}^m l_{1,i} \int_{z_i} \{g(y_1 + z_i, y_0, v_1, l_1 + \beta_i, v_0, l_0, \delta) - g(y_1, y_0, v_1, l_1, v_0, l_0, \delta)\} f_{Z_i}(z_i) dz_i, \end{aligned} \quad (12)$$

where  $\mu^X$  is the vector of expected log-returns,  $\mu^V$  the vector of stochastic volatility drifts,  $\mu^\lambda$  the vector of jump intensity drift,  $\sigma^{YY}$  the variance-covariance matrix of log-returns,  $\sigma^{YV}$  the variance-covariance matrix of interactions between returns and volatilities,  $\sigma^{VV}$  the variance-covariance matrix

of stochastic volatilities and finally  $\beta_i = [\beta_{i,1}, \dots, \beta_{i,m}]'$  the vector of excitation parameters for the  $i^{\text{th}}$  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

$$\begin{aligned} \mathbb{E}_{Y_1, \xi_1} [g(Y_1, Y_0, \xi_1, \xi_0, \Delta) | Y_0, \xi_0] &= \exp(\Delta \mathcal{A}) \cdot g(Y_0, Y_0, \xi_0, \xi_0, 0) \\ &= \sum_{j=0}^J \frac{\Delta^j}{j!} (\mathcal{A}^j \cdot g)(Y_0, Y_0, \xi_0, \xi_0, 0) + O_p(\Delta^{J+1}), \end{aligned} \quad (13)$$

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, (12) and its iterates  $\mathcal{A}^j \cdot g$ , hence the terms in (13), can be evaluated in closed-form for the moment functions of interest given in (11). While the starting moment function  $f(y_1, y_0, \delta)$  does not depend directly upon the latent variables  $(\xi_1, \xi_0)$ ,  $\mathcal{A} \cdot f$  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,  $\mathbb{E}_{Y_1, \xi_1} [g(Y_1, Y_0, \xi_1, \xi_0, \Delta) | Y_0, \xi_0]$ . 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

$$\begin{aligned} \mathbb{E}_{Y_1, Y_0, \xi_1, \xi_0} [g(Y_1, Y_0, \xi_1, \xi_0, \Delta)] &= \mathbb{E}_{Y_1, Y_0, \xi_1, \xi_0} [\mathbb{E}_{Y_1, \xi_1} [g(Y_1, Y_0, \xi_1, \xi_0, \Delta) | Y_0, \xi_0]] \\ &= \sum_{j=0}^J \frac{\Delta^j}{j!} \mathbb{E}_{Y_0, \xi_0} [(\mathcal{A}^j \cdot g)(Y_0, Y_0, \xi_0, \xi_0, 0)] + O_p(\Delta^{J+1}), \end{aligned} \quad (14)$$

so the last step in the necessary calculations will involve computing unconditional expectations with respect to the stationary law of the state variables. We will provide the expressions for these moment functions in the theorems that follow.

### 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 for illustration purposes 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:

$$\begin{cases} dX_t = \mu dt + \sqrt{V_t} dW_t^X + Z_t dN_t \\ dV_t = \kappa(\theta - V_t)dt + \eta\sqrt{V_t} dW_t^V \\ d\lambda_t = \alpha(\lambda_\infty - \lambda_t)dt + \beta dN_t \end{cases} \quad (15)$$

with  $\mathbb{E}[dW_t^X dW_t^V] =: \rho^V dt$  and  $\lambda := \mathbb{E}[\lambda_t] = \alpha\lambda_\infty/(\alpha - \beta)$ . 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  $(N, \lambda)$  which facilitates the analysis of the model. But it is possible to introduce a level effect and an asymmetry effect by replacing the term  $\beta dN_t$  in the third equation of (15) with a term  $\phi(Z_t)dN_t$  where  $\phi$  is a function selected jointly with the distribution of  $Z_t$  to maintain the positivity and stationarity of  $\lambda_t$ . 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  $M[Z, k] := \mathbb{E}[Z_t^k]$ . Our first main result gives the moment functions explicitly:

**Theorem 1** *For the univariate model (15), the moments are given in closed-form up to order  $\Delta^2$  by the following expressions*

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

while the autocovariance function of log-returns is given by

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

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.<sup>9</sup> The autocovariances isolate the jump parameters. Indeed, if the model had no jump component, then from (15) with  $Z_t \equiv 0$ , the law of iterated expectations implies that

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

and so  $\mathbb{E}[(\Delta X_t - \mathbb{E}[\Delta X_t])(\Delta X_{t+\tau} - \mathbb{E}[\Delta X_{t+\tau}])] = 0$ . Thus any autocovariance is due to the jump component. Further, if the jump component were 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 autocovariances of the increments isolate the self-exciting component of the model. Note that jumps and volatility are separated by the higher order moments (3 and 4), whether volatility is stochastic or not. Additionally, but only if volatility is not stochastic, the autocovariance function of *squared returns* also identifies the jump component:

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

for all  $\tau > 0$ ,  $r = 1, 2$ .

As noted above, the model reduces to a Poissonian jump-diffusion if  $\beta = 0$ . Then we indeed have

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

and the model does no longer generate any autocorrelation in returns.

We note that if the expected jump size is exactly 0 ( $M[Z, 1] = 0$ ) the autocovariance function of returns becomes identical zero, but the self-exciting component of the model may still be identified by the autocovariance function of squared returns. The autocovariance function of squared returns may

---

<sup>9</sup>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.

therefore best be viewed as the jump analog of an ARCH-type moment function for volatility. Finally, we note that the expressions above are given in closed-form as Taylor expressions in the sampling interval  $\Delta$  truncated after the terms of order  $\Delta^2$ . In practice, for (bi)daily returns ( $\Delta = 1/252$  or  $\Delta = 1/504$ ), the approximation error is very small; theoretically, any truncated expression leads to an asymptotic bias term which is in practice negligible compared to the sampling variation of the estimates but that, if necessary, can be eliminated theoretically by assuming infill asymptotics.

### 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

$$\begin{cases} dX_{1t} = \mu_1 dt + \sigma_1 dW_t^1 + Z_{1t} dN_{1t} \\ dX_{2t} = \mu_2 dt + \sigma_2 dW_t^2 + Z_{2t} dN_{2t} \\ d\lambda_{1t} = \alpha_1 (\lambda_{1\infty} - \lambda_{1t}) dt + \beta_{11} dN_{1t} \\ d\lambda_{2t} = \alpha_2 (\lambda_{2\infty} - \lambda_{2t}) dt + \beta_{21} dN_{1t} + \beta_{22} dN_{2t} \end{cases} \quad (16)$$

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

**Theorem 2** *For the bivariate model (16), the mean and variance of the log-returns are given in closed-form up to order  $\Delta^2$  by the following expressions*

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

and

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

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  $\Delta^2$  by the following expressions:*

$$\begin{aligned}
\mathbb{E}[(\Delta X_{1t} - \mathbb{E}[\Delta X_{1t}])^3] &= \lambda_1 M[Z_1, 3] \Delta + \frac{3(2\alpha_1 - \beta_{11})\beta_{11}\lambda_1}{2(\alpha_1 - \beta_{11})} M[Z_1, 1] M[Z_1, 2] \Delta^2 + o(\Delta^2) \\
\mathbb{E}[(\Delta X_{2t} - \mathbb{E}[\Delta X_{2t}])^3] &= \lambda_2 M[Z_2, 3] \Delta + \frac{3}{2} \left( \frac{(2\alpha_2 - \beta_{22})\beta_{22}\lambda_2}{(\alpha_2 - \beta_{22})} \right. \\
&\quad \left. + \frac{(\alpha_1^2 + (\alpha_1 - \beta_{11})(\alpha_2 - \beta_{22}))\lambda_1\beta_{21}^2}{(\alpha_1 - \beta_{11})(\alpha_2 - \beta_{22})(\alpha_1 + \alpha_2 - \beta_{11} - \beta_{22})} \right) M[Z_2, 1] M[Z_2, 2] \Delta^2 + o(\Delta^2) \\
\mathbb{E}[(\Delta X_{1t} - \mathbb{E}[\Delta X_{1t}])^2(\Delta X_{2t} - \mathbb{E}[\Delta X_{2t}])] &= \frac{\beta_{21}(\alpha_1^2 + (\alpha_2 - \beta_{22})(\alpha_1 - \beta_{11}))\lambda_1}{2(\alpha_1 - \beta_{11})(\alpha_1 + \alpha_2 - \beta_{11} - \beta_{22})} \\
&\quad \times M[Z_1, 2] M[Z_2, 1] \Delta^2 + o(\Delta^2) \\
\mathbb{E}[(\Delta X_{1t} - \mathbb{E}[\Delta X_{1t}])^2(\Delta X_{2t} - \mathbb{E}[\Delta X_{2t}])^2] &= \frac{\beta_{21}(\alpha_1^2 + (\alpha_2 - \beta_{22})(\alpha_1 - \beta_{11}))\lambda_1}{2(\alpha_1 - \beta_{11})(\alpha_1 + \alpha_2 - \beta_{11} - \beta_{22})} \\
&\quad \times M[Z_1, 1] M[Z_2, 2] \Delta^2 + o(\Delta^2) \\
\mathbb{E}[(\Delta X_{1t} - \mathbb{E}[\Delta X_{1t}])^4] &= \lambda_1 M[Z_1, 4] \Delta + (3\sigma_1^4 + 6M[Z_1, 2] \lambda_1 \sigma_1^2 \\
&\quad + \frac{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})} + \frac{2(2\alpha_1 - \beta_{11})\beta_{11}\lambda_1}{(\alpha_1 - \beta_{11})} M[Z_1, 1] M[Z_1, 3]) \Delta^2 + o(\Delta^2) \\
\mathbb{E}[(\Delta X_{2t} - \mathbb{E}[\Delta X_{2t}])^4] &= \lambda_2 M[Z_2, 4] \Delta + (3\sigma_2^4 + 6M[Z_2, 2] \lambda_2 \sigma_2^2 \\
&\quad + 3M[Z_2, 2]^2 (\eta_{22} + \beta_{22}\lambda_2) + 4(\eta_{22} + (\beta_{22} - \lambda_2)\lambda_2) M[Z_2, 1] M[Z_2, 3]) \Delta^2 + o(\Delta^2).
\end{aligned}$$

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

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

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

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

and

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

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

**Theorem 4** *For the bivariate model (15), the autocovariances of the log-returns and squared log-returns at lag  $\tau > 0$ ,*

$$C_{i,j}^r(\tau) := \mathbb{E}[(\Delta X_{j,t})^r - \mathbb{E}[(\Delta X_{j,t})^r])(\Delta X_{i,t+\tau})^r - \mathbb{E}[(\Delta X_{i,t+\tau})^r]], \quad r = 1, 2$$

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

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

where

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

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 ((bi)daily). However, in order to improve accuracy and also limit the number of moment conditions required over longer time intervals, we calculate in addition the unconditional interval-based (that is, exact rather than approximated in  $\Delta$ ) moments ( $i, j = 1, 2; 0 \leq s_1 < s_2 < s_3 < s_4$ )

$$\left\{ \begin{array}{l} \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{i,t} \right] \\ \mathbb{E} \left[ \left( \int_{s_1}^{s_2} dX_{i,t} - \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{i,t} \right] \right)^2 \right] \\ \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{i,t} \int_{s_1}^{s_2} dX_{j,u} - \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{i,t} \right] \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{j,u} \right] \right], \quad i \neq j \\ \mathbb{E} \left[ \int_{s_3}^{s_4} (dX_{i,u})^r \int_{s_1}^{s_2} (dX_{j,t})^r - \mathbb{E} \left[ \int_{s_3}^{s_4} (dX_{i,u})^r \right] \mathbb{E} \left[ \int_{s_1}^{s_2} (dX_{j,t})^r \right] \right], \quad r = 1, 2, \end{array} \right.$$

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  $\Omega$ ; see (20) 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  $Y_{n\Delta} := X_{n\Delta} - X_{(n-1)\Delta}$ ,  $n = 1, \dots, N$ ,  $\Delta > 0$ , with  $N\Delta = T$ , denote the log-returns on the interval  $[0, T]$  and let  $\vartheta \in \Theta$  denote our  $d$ -dimensional parameter vector. To estimate  $\vartheta$  we consider a vector of  $M$  moment conditions  $h(y, \Delta, \vartheta)$ ,  $M \geq d$ , continuously differentiable in  $\vartheta$ . Let  $\vartheta_0$  denote the true value of  $\vartheta$  and suppose that  $\mathbb{E}[h(Y_{n\Delta}, \Delta, \vartheta_0)] = 0$ . 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  $g_T(y, \Delta, \vartheta)$  denote the sample average of  $h(y_{n\Delta}, \Delta, \vartheta)$ , that is  $g_T(y, \Delta, \vartheta) := \frac{1}{N} \sum_{n=1}^N h(y_{n\Delta}, \Delta, \vartheta)$ . Then the GMM estimator  $\hat{\vartheta}_T$  is the value of  $\vartheta \in \Theta$  that minimizes the quadratic form  $g_T(y, \Delta, \vartheta)' W_T g_T(y, \Delta, \vartheta)$ , 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  $\{h(Y_{n\Delta}, \Delta, \vartheta_0)\}_{n=-\infty}^{\infty}$  is strictly stationary with mean zero and  $v$ -th autocovariance matrix  $\Gamma_v := \mathbb{E}[h(Y_{n\Delta}, \Delta, \vartheta_0) h(Y_{(n-v)\Delta}, \Delta, \vartheta_0)']$ . 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  $h(y_{n\Delta}, \Delta, \vartheta_0)$ :

$$S = \lim_{N \rightarrow \infty} N \mathbb{E}[g_T(Y, \Delta, \vartheta_0) g_T(Y, \Delta, \vartheta_0)']. \quad (17)$$

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

$$\hat{S}_T = \hat{\Gamma}_{0,T} + \sum_{v=1}^q \left(1 - \frac{v}{q+1}\right) (\hat{\Gamma}_{v,T} + \hat{\Gamma}_{v,T}'), \quad (18)$$

with

$$\hat{\Gamma}_{v,T} = \frac{1}{N} \sum_{n=v+1}^N h(y_{n\Delta}, \Delta, \tilde{\vartheta}) h(y_{(n-v)\Delta}, \Delta, \tilde{\vartheta})'. \quad (19)$$

Here  $\tilde{\vartheta}$  is an initial consistent estimate of  $\vartheta_0$ , which can be obtained by minimizing the quadratic

form with  $W_T = I$ . Under standard regularity conditions (see Hansen (1982)), as  $T \rightarrow \infty$ ,  $\hat{\vartheta}$  is asymptotically normal with asymptotic variance  $\Omega$  where

$$\Omega^{-1} := \Delta^{-1} D' W D (D' W S W D)^{-1} D' W D, \quad (20)$$

and  $D$  is the gradient of  $E[h(Y_{n\Delta}, \Delta, \vartheta)]$  with respect to  $\vartheta'$  evaluated at  $\vartheta'_0$ . When  $W_T$  is chosen optimally,  $W = S^{-1}$  and (20) reduces to  $\Omega^{-1} = \Delta^{-1} D' S^{-1} D$ .

Note that the matrices  $D$  and  $S$  depend on  $\Delta$  so that if high frequency asymptotics were used the rate of convergence of each parameter would not necessarily be  $\sqrt{T}$  but would depend on  $\Delta$  as well. Also, as the moments are given by Taylor expansions in  $\Delta$ , a bias term of small order in  $\Delta$  is left, an effect similar to that in Aït-Sahalia and Mykland (2003). We treat this term as negligible on the basis of (bi)daily data. We note that the parameters of the model can be fully identified only if  $\lambda_\infty > 0$ , which we will maintain throughout. And a test of  $\lambda_\infty > 0$  against a null of  $\lambda_\infty = 0$  cannot be performed via a usual  $F$ -statistic or Wald test, because of the presence of nuisance parameters not identified under the null.

### 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, 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. The Monte Carlo results are not included in the paper to save space. They are available from the authors upon request.<sup>10</sup>

In general, 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 (exact i.e., interval-based), daily instantaneous (co)variances (exact) and daily higher order moments (3 and 4; approximated at

---

<sup>10</sup>From the onset, one should be aware of the fact that one cannot expect degrees of accuracy similar or even close to what we are used to when estimating, e.g., a standard stochastic volatility model. By definition, extreme events occur infrequently and there is a positive probability that no jump occurs in any given finite time interval. When no jump occurs, there is no identification whatsoever, while when few jumps occur identification is expected to be weak. This is a consequence of the classical peso problem. Furthermore, we are effectively after subtle effects concerning the jumps of the process, not just whether they are present or not, but their finer structure: whether they self-excite, whether they cross-excite, etc. Finally, the estimation procedure can rely only on the time series of the log-returns of the various assets, since all other state variables are latent (volatility, jumps, jump intensities). In other words, we are asking for a lot out of the data. So one should not expect miracles from a time series of necessarily finite length. Nevertheless, what emerges out of the Monte Carlos is evidence that, using our estimation methodology, the parameters of the data generating process can be recovered with sufficient degree of precision in a realistic context.

the leading order in  $\Delta$ ) and to use daily, weekly or monthly autocovariances and cross-covariances (both exact) 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.

There are two latent processes,  $V_t$  and  $\lambda_t$ ; while those are integrated out as part of the development of the unconditional moments, their corresponding parameters remain. So while they are theoretically identified, 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, and identify only the level of the volatility. In the bivariate case, the restrictions that  $\alpha_1 = \alpha_2$  and  $\lambda_{1,\infty} = \lambda_{2,\infty}$  facilitate the identification. Since our focus is on estimating the parameters of the mutually exciting process, we set  $\gamma_{i,1} = \gamma_{i,2} =: \gamma_i$  and pre-identify  $\gamma_i$  by considering the average of absolute returns that exceed 2%. Furthermore, we pre-identify  $p_i$  from the ratio of the third and fourth moments that are approximated at the leading order in  $\Delta$ , which depends only on  $p_i$ , given  $\gamma_i$ . Pre-identified parameters are treated as fixed and given.

To find the required starting values to be used in the GMM maximization, we adopt a natural two-stage procedure that helps disentangling diffusion from jumps: we first estimate the parameters  $\{\mu_1, \mu_2, \theta_1, \theta_2, \rho\}$  from the continuous part of our model using truncated data. We fit the truncated data using moment conditions that are derived from the continuous part of our model, ignoring the discontinuous jump part. 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  $\{\lambda_\infty, \alpha, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$  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.<sup>11</sup> 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

---

<sup>11</sup>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. Due to the truncation, the *preliminary* volatility and correlation parameters thus obtained are slightly biased downwards, which we eventually correct for. Indeed, once the starting values have obtained, the full set of moment conditions is applied to the original non-truncated data. Doing so removes the slight downward bias in the preliminary correlation and volatility parameters. We repeat this procedure 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.

Carlo simulated data are usually quite accurate. It behaves as is typical of GMM given the number of parameters of the model. Higher-dimensional ( $m > 2$ ) versions of the model become more difficult to estimate. 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, which we don't). Another, which we have primarily adopted, consists in focusing only on fairly unrestricted but *bivariate* models. The Monte Carlo results show that, using our estimation methodology, 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 have 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  $\beta_{22}$  that are not significantly different from zero in this case.

## 4. Empirical Analysis

### 4.1. Data and Data Sequencing

We are interested in capturing patterns of contagion between stock markets around the world. 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 and which would much be required when plain daily data were used for the analysis, we use open-to-close (= “daylight” return) and close-to-open returns (= “night” return). Such data have also been used e.g., by Becker et al. (1990) and Hamao et al. (1990) to study correlation in international equity markets.

We collect open and close international equity index data from [finance.yahoo.com](http://finance.yahoo.com). We will study five indices: S&P 500 (US); FTSE 100 (UK); Nikkei 225 (Japan, JA); Hang Seng (Hong Kong, HK); and IPC (Mexico, MX). Open and close data start on: January 2, 1980 (for US), April 2, 1984 (for UK), January 4, 1984 (for JA), December 31, 1986 (for HK), and November 8, 1991 (for MX). For the US, we also consider the subsample that starts on January 2, 1990, referred to as US90; we sometimes use the acronym US80 to refer to the full sample. For each pair, the (joint) sample period is dictated by the market with the shortest sample period. Our sample period ends on April 30, 2013, for all indices. Summary statistics are in Table 1. (Note that the summary statistics are computed over the different individual sample periods.)

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 US80 return series

and the US90 return series. 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.

The model predicts that the empirical autocorrelation and cross-correlation is decreasing and convex, hence that the instantaneous correlation is the largest. Different time zones and hours of operation of the markets around the world require careful treatment to get the sequencing right given the time differences.<sup>12</sup> Our treatment of the data is illustrated in Figure 4. For the US and Japan, we define as instantaneous the overnight return in Japan and the corresponding day return in the US; and the day return in Japan and the corresponding overnight return in the US. The same applies for the pair US and Hong Kong. Similarly, for the US and UK, we define as instantaneous the overnight return in the UK and the corresponding day return in the US; and the day return in the UK and the corresponding overnight return in the US. Note that no information (“correlation”) is lost in this treatment of the data. We finally note that the pair US and Mexico does not require any modification, because these markets operate in the same time zone, and is treated as synchronous.

Care is required when matching the time series as closing days are not the same across different markets. When one market is closed, say Japan, we wait in both markets until the next re-open (of Japan, in this case) and make all the time until the next re-open matched to the previous time period (the one containing, in this case, the Japanese market’s close).

As an example of the time series data thus obtained, we plot the autocovariances and cross-covariances of squared returns for the pair (US, JA) in Figure 5. The data reveals substantial autocorrelation between US squared returns on day part  $t$  and US squared returns on the following day part  $t + 2$  (see Figure 4 for the day parts definitions). We also observe cross-correlation between US squared returns on day part  $t$  and squared returns of Japan on the following day part  $t + 1$ . The presence of these non-zero correlations is what makes possible the identification of the self- and cross-excitation parameters.

---

<sup>12</sup>To get a qualitative insight in the direction of jump transmissions, we sorted daily US returns to find the most extreme declines (over 3.0% in a single day) in the US stock market. 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 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 (or recently, PIGS) markets, and then other regions of the world decline once the US start declining. There are few exceptions to this general pattern of transmission. The detailed results of this qualitative analysis are available from the authors upon request.

## 4.2. Parameter Estimates

### 4.2.1 The Univariate Model

We consider the univariate case first. We use  $[4 + (\text{\#lags of autocovariances})]$  moment conditions to identify the parameters of the univariate version of our Hawkes jump-diffusion model: the interval-based first moments and autocovariances of squared returns, and the third and fourth moments approximated at the leading order in  $\Delta$ . This leaves us with 5 parameters to be identified:  $\alpha, \beta, \lambda_\infty, \theta, \mu$ .

Visual inspection of the autocorrelograms and (in the bivariate case below) 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 eventually build a limited number of interval-based autocovariances over a 50 bidaily lags time period. GMM parameter estimates are in Table 2.

We find large estimated values for  $\alpha$  and  $\beta$ , providing clear evidence of self-excitation, both in the US market as well as in all the other markets. These estimates come along with relatively moderate estimates for the volatility levels. It appears that, thanks 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. The model implied moments computed with the obtained parameter estimates fit the data remarkably well.

### 4.2.2 The Bivariate Model

Next, we consider the bivariate version of the Hawkes jump-diffusion model. 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 of squared returns, and the third and fourth moments approximated at the leading order in  $\Delta$ . Informed by Monte Carlo evidence, we restrict  $\alpha_1 = \alpha_2 =: \alpha$  and  $\lambda_{1,\infty} = \lambda_{2,\infty} =: \lambda_\infty$ . This leaves us with 11 parameters to be identified:  $\alpha, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \lambda_\infty, \theta_1, \theta_2, \rho, \mu_1, \mu_2$ . As in the univariate case, we build a limited number of interval-based autocovariances and cross-covariances over a 50 bidaily lags time period. GMM parameter estimates are in Table 3. One round of optimization (data import, GMM optimization, standard errors and test statistics) of the model takes only a few minutes on a regular computer.

We find large estimated values for  $\beta_{11}$  and  $\beta_{22}$ , 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 again come along with relatively moderate estimates for the volatility levels, as in the univariate case. Also, we typically find relatively large estimated values for  $\beta_{21}$ , measuring the degree of transmission from the US to other regions of the world. Large values for  $\beta_{21}$  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 empirical cross-correlation plots, the effect seems to be mainly driven by transmission on the first few day parts following the day of occurrence of a US jump. Estimates of  $\beta_{12}$ , measuring the degree of the reverse transmission, are relatively small. Statistically, the estimated values of the  $\beta_{12}$ 's are not significantly different from zero. 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.3), (B.4) and (B.5), as well as the reconstructed third and fourth moments as a means to validate the empirical fit of the model.

#### 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 et al. (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  $\beta$  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 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  $\chi^2$  statistic that follows from (20). 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:  $\mathcal{H}_0^I : \beta_{i,j} = 0, i, j = 1, 2$ ;  $\mathcal{H}_0^{II} : \beta_{i,i} = 0, i = 1, 2$ ;  $\mathcal{H}_0^{III} : \beta_{i,j} = 0, i, j = 1, 2, i \neq j$ . We adopt a Wald chi-square test based on the GMM estimates. Test results are reported in Table 4. The null hypotheses are rejected in the vast majority of the cases, providing clear evidence for excitation (rejection of  $\mathcal{H}_0^I$ ), self-excitation (rejection of  $\mathcal{H}_0^{II}$ ) and cross-excitation (rejection of  $\mathcal{H}_0^{III}$ ). It means in particular that a Poisson jump-diffusion model is rejected when tested against the Hawkes jump-diffusion model.

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 et al. (2010)).

#### 4.4. *Robustness Check: Sample Period*

As a robustness check, we have estimated the Hawkes jump-diffusion model over a subsample 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 shrunk the US sample to cover January 2, 1990, to April 30, 2013, which corresponds roughly to two thirds of the sample previously employed ( $[1/3, 1]$ ). The parameter estimates we find are qualitatively consistent with our earlier findings, see Table 2: there is clear evidence for self-excitation in the US market. We find that the expected number of jumps in the subsample is estimated at smaller values than the expected number of jumps in the original sample. This is largely due to the fact that the original sample covers the crisis episode in 1987, while this episode is not included in the subsample.

#### 4.5. *Measuring Market Stress Using Filtered Values of the Jump Intensities*

Stress in the financial 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 instead

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.

We therefore propose to use the jump intensities from the model as an alternative 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 our Hawkes jump-diffusion 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 can employ the parameters that were just estimated: insert the results of Table 3 in (8), and estimate to be a jump (hence,  $dN_{j,t} = 1$ ) each return in the corresponding time series that exceeds 2% in absolute value. Filtering jumps on the basis of large returns is a natural approach, see e.g., Lee and Mykland (2008). Adaptive cutoffs based on a preliminary estimate of the volatility component of the model lead to similar empirical results.

The resulting plots of the estimated time series of  $(\lambda_{US,t}, \lambda_{UK,t})$  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), the fall of 2008 (credit crisis) and the 2011-12 period (Eurozone crisis). We note that, as implied by the model, jump intensities in both markets jump together, but price jumps themselves are not necessarily synchronous.

Figure 7, based on our model, can be thought of as an alternative measure of market stress to that based on VIX in Figure 6, with the Brownian volatility removed in Figure 7 from the total quadratic variation in VIX to focus exclusively on the risk that is potentially the most damaging, that due to jumps. As a result, periods of crisis are more apparent in Figure 7, with sharp increases in jump intensities, whereas these periods are somewhat obscured in Figure 6 due to the inclusion of the continuous volatility in VIX, which effectively acts as noise that masks the jumps.

#### 4.6. Out-of-Sample Predictions

Once a first jump is identified in the time series, the model predicts that further jumps are more likely to occur in the short run. We now 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 market. Specifically, we define a jump indicator taking the value one if the US index return exceeds 2% in absolute value on a full trading day, 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  $\Delta = 1$  (full) 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 exceeding 2% in absolute value, given a realized volatility estimate (based on the full history or on a rolling window). The predicted probabilities of the occurrence of an extreme return (“jump”) over a full trading day are plotted in Figure 8 for the different models. On the basis of the root Mean Squared Prediction Error as a standard metric to compare the forecast performance of the different models, measuring the average squared difference between the predicted jump probability and the jump indicator, we find that the Hawkes jump-diffusion model clearly dominates the nested models with simpler jumps or without jumps.

## 5. Conclusions

We have 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, in line with the behavior of the data. Unlike models in the literature based on Hawkes pure jumps, our *Hawkes jump-diffusion 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 have derived closed-form moment functions which make estimation and testing of the model feasible.

We have implemented the estimation procedure on international equity data, studying the patterns of jump excitation among five world markets. Our empirical results indicate that both the US market as well as the other markets strongly self-excited during the recent financial crisis. We find that US jumps tend to get reflected quickly in most other markets, while statistical evidence for the reverse transmission is much less pronounced. The tests empirically reject a jump-diffusion model based on Poissonian jumps against a similar model based on Hawkes jumps. Filtered jump-intensities can act as a measure of market stress, with good out-of-sample properties. Finally, the general model can be restricted to be affine, which opens the door to potential further applications such as derivative pricing and optimal portfolio selection.

## Appendix: Explicit Formulae for the Moment Conditions

### Appendix A: Univariate Self-Exciting Jumps

Let  $m = 1$  and consider the model (15) with  $\mathbb{E}[dW_t^X dW_t^V] =: \rho^V dt$ . The corresponding integral equation for  $\lambda_t$  reads:

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

Write

$$\lambda := \mathbb{E}[\lambda_t], \quad V_N(\tau) := \mathbb{E}[dN_{t+\tau} dN_t] / (dt)^2 - \lambda^2.$$

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

$$\frac{\mathbb{E}[dX_{t+\tau} dX_t]}{(dt)^2} - \left( \frac{\mathbb{E}[dX_t]}{dt} \right)^2 = (\mathbb{E}[Z])^2 V_N(\tau), \quad \tau > 0; \quad (\text{A.1})$$

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,

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

We now turn to interval-based moment conditions. 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  $\lambda := \mathbb{E}[\lambda_t]$ , and

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

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

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

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,

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

with

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

Note that if volatility is not stochastic,

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

With stochastic volatility of the Heston type, the above autocovariance function of squared log-returns of the self-exciting jump model will simply convolute with the well-known autocovariance function of squared log-returns of the pure Heston model.

## Appendix B: Bivariate Mutually Exciting Jumps

Let  $m = 2$  and assume that

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

with  $\mathbb{E}[dW_{1,t}^X dW_{2,t}^X] =: \rho dt$  and  $\mathbb{E}[dW_{i,t}^X dW_t^V] =: \rho_i^V dt$ ,  $i = 1, 2$ . 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:

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

Write

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

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

$$\begin{pmatrix} \frac{\mathbb{E}[dX_{1,t+\tau} dX_{1,t}]}{(dt)^2} - \left(\frac{\mathbb{E}[dX_{1,t}]}{dt}\right)^2 & \frac{\mathbb{E}[dX_{1,t+\tau} dX_{2,t}]}{(dt)^2} - \frac{\mathbb{E}[dX_{1,t}]}{dt} \frac{\mathbb{E}[dX_{2,t}]}{dt} \\ \frac{\mathbb{E}[dX_{2,t+\tau} dX_{1,t}]}{(dt)^2} - \frac{\mathbb{E}[dX_{1,t}]}{dt} \frac{\mathbb{E}[dX_{2,t}]}{dt} & \frac{\mathbb{E}[dX_{2,t+\tau} dX_{2,t}]}{(dt)^2} - \left(\frac{\mathbb{E}[dX_{2,t}]}{dt}\right)^2 \end{pmatrix} \quad (B.2)$$

$$= \begin{pmatrix} (\mathbb{E}[Z_1])^2 V_{1,1,N}(\tau) & \mathbb{E}[Z_1] \mathbb{E}[Z_2] V_{1,2,N}(\tau) \\ \mathbb{E}[Z_1] \mathbb{E}[Z_2] V_{2,1,N}(\tau) & (\mathbb{E}[Z_2])^2 V_{2,2,N}(\tau) \end{pmatrix}, \quad \tau > 0;$$

with

$$\lambda_1 = \frac{\lambda_{1,\infty} \alpha_1 (\alpha_2 - \beta_{22}) + \lambda_{2,\infty} \alpha_2 \beta_{12}}{(\alpha_1 - \beta_{11})(\alpha_2 - \beta_{22}) - \beta_{12} \beta_{21}},$$

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

and

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

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

where

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

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,

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

The interval-based moment conditions are as follows. 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

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

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

$$\begin{aligned}
& \left( \frac{\mathbb{E} \left[ \int_{s_3}^{s_4} dX_{1,u} \int_{s_1}^{s_2} dX_{1,t} \right]}{\Delta_1 \Delta_2} - \frac{\mathbb{E} \left[ \int_{s_3}^{s_4} dX_{1,u} \right] \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{1,t} \right]}{\Delta_2 \Delta_1} \quad \frac{\mathbb{E} \left[ \int_{s_3}^{s_4} dX_{1,u} \int_{s_1}^{s_2} dX_{2,t} \right]}{\Delta_1 \Delta_2} - \frac{\mathbb{E} \left[ \int_{s_3}^{s_4} dX_{1,u} \right] \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{2,t} \right]}{\Delta_2 \Delta_1} \right) \\
& \left( \frac{\mathbb{E} \left[ \int_{s_3}^{s_4} dX_{2,u} \int_{s_1}^{s_2} dX_{1,t} \right]}{\Delta_1 \Delta_2} - \frac{\mathbb{E} \left[ \int_{s_3}^{s_4} dX_{2,u} \right] \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{1,t} \right]}{\Delta_2 \Delta_1} \quad \frac{\mathbb{E} \left[ \int_{s_3}^{s_4} dX_{2,u} \int_{s_1}^{s_2} dX_{2,t} \right]}{\Delta_1 \Delta_2} - \frac{\mathbb{E} \left[ \int_{s_3}^{s_4} dX_{2,u} \right] \mathbb{E} \left[ \int_{s_1}^{s_2} dX_{2,t} \right]}{\Delta_2 \Delta_1} \right) \\
& = \begin{pmatrix} \frac{(\mathbb{E}[Z_1])^2 I_{1,1,N}^\tau(\tau, \Delta_1, \Delta_2)}{\Delta_1 \Delta_2} & \frac{\mathbb{E}[Z_1] \mathbb{E}[Z_2] I_{1,2,N}^\tau(\tau, \Delta_1, \Delta_2)}{\Delta_1 \Delta_2} \\ \frac{\mathbb{E}[Z_1] \mathbb{E}[Z_2] I_{2,1,N}^\tau(\tau, \Delta_1, \Delta_2)}{\Delta_1 \Delta_2} & \frac{(\mathbb{E}[Z_2])^2 I_{2,2,N}^\tau(\tau, \Delta_1, \Delta_2)}{\Delta_1 \Delta_2} \end{pmatrix};
\end{aligned} \tag{B.3}$$

with

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

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

Appendix B. Furthermore,

$$\frac{\mathbb{E}\left[\left(\int_{s_1}^{s_2} dX_{i,t}\right)^2\right]}{\Delta_1} - \frac{\left(\mathbb{E}\left[\int_{s_1}^{s_2} dX_{i,t}\right]\right)^2}{\Delta_1} = \theta_i + \lambda_i \mathbb{E}\left[Z_i^2\right] + \frac{2(\mathbb{E}[Z_i])^2 I_{i,i,N}(\Delta_1)}{\Delta_1}, \quad i = 1, 2; \quad (\text{B.4})$$

$$\frac{\mathbb{E}\left[\int_{s_1}^{s_2} dX_{1,t} \int_{s_1}^{s_2} dX_{2,u}\right]}{\Delta_1} - \frac{\mathbb{E}\left[\int_{s_1}^{s_2} dX_{1,t}\right] \mathbb{E}\left[\int_{s_1}^{s_2} dX_{2,u}\right]}{\Delta_1} = \rho \sqrt{\theta_1 \theta_2} + \frac{\mathbb{E}[Z_1] \mathbb{E}[Z_2] (I_{1,2,N}(\Delta_1) + I_{2,1,N}(\Delta_1))}{\Delta_1}; \quad (\text{B.5})$$

with

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

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

Note that if volatility is not stochastic,

$$\left( \begin{array}{c} \frac{\mathbb{E}\left[\int_{s_3}^{s_4} (dX_{1,u})^k \int_{s_1}^{s_2} (dX_{1,t})^k\right]}{\Delta_1 \Delta_2} - \frac{\mathbb{E}\left[\int_{s_3}^{s_4} (dX_{1,u})^k\right] \mathbb{E}\left[\int_{s_1}^{s_2} (dX_{1,t})^k\right]}{\Delta_2 \Delta_1} \\ \frac{\mathbb{E}\left[\int_{s_3}^{s_4} (dX_{2,u})^k \int_{s_1}^{s_2} (dX_{1,t})^k\right]}{\Delta_1 \Delta_2} - \frac{\mathbb{E}\left[\int_{s_3}^{s_4} (dX_{2,u})^k\right] \mathbb{E}\left[\int_{s_1}^{s_2} (dX_{1,t})^k\right]}{\Delta_2 \Delta_1} \\ \frac{\mathbb{E}\left[\int_{s_3}^{s_4} (dX_{1,u})^k \int_{s_1}^{s_2} (dX_{2,t})^k\right]}{\Delta_1 \Delta_2} - \frac{\mathbb{E}\left[\int_{s_3}^{s_4} (dX_{1,u})^k\right] \mathbb{E}\left[\int_{s_1}^{s_2} (dX_{2,t})^k\right]}{\Delta_2 \Delta_1} \\ \frac{\mathbb{E}\left[\int_{s_3}^{s_4} (dX_{2,u})^k \int_{s_1}^{s_2} (dX_{2,t})^k\right]}{\Delta_1 \Delta_2} - \frac{\mathbb{E}\left[\int_{s_3}^{s_4} (dX_{2,u})^k\right] \mathbb{E}\left[\int_{s_1}^{s_2} (dX_{2,t})^k\right]}{\Delta_2 \Delta_1} \end{array} \right) \\ = \left( \begin{array}{c} \frac{(\mathbb{E}[Z_1^k])^2 I_{1,1,N}^\tau(\tau, \Delta_1, \Delta_2)}{\Delta_1 \Delta_2} \\ \frac{\mathbb{E}[Z_1^k] \mathbb{E}[Z_2^k] I_{1,2,N}^\tau(\tau, \Delta_1, \Delta_2)}{\Delta_1 \Delta_2} \\ \frac{\mathbb{E}[Z_1^k] \mathbb{E}[Z_2^k] I_{2,1,N}^\tau(\tau, \Delta_1, \Delta_2)}{\Delta_1 \Delta_2} \\ \frac{(\mathbb{E}[Z_2^k])^2 I_{2,2,N}^\tau(\tau, \Delta_1, \Delta_2)}{\Delta_1 \Delta_2} \end{array} \right), \quad k = 1, 2.$$

With stochastic volatility of the Heston type, the above autocovariance and cross-covariance functions of squared log-returns ( $k = 2$ ) of the mutually exciting jump model will simply convolute with the well-known autocovariance and cross-covariance functions of squared log-returns of the pure Heston model.



## References

- Aït-Sahalia, Y., 2002. Telling from discrete data whether the underlying continuous-time model is a diffusion. *The Journal of Finance* 57, 2075–2112.
- Aït-Sahalia, Y., 2004. Disentangling diffusion from jumps. *Journal of Financial Economics* 74, 487–528.
- Aït-Sahalia, Y., Jacod, J., 2009. Testing for jumps in a discretely observed process. *Annals of Statistics* 37, 184–222.
- Aït-Sahalia, Y., Mykland, P. A., 2003. The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* 71, 483–549.
- Barndorff-Nielsen, O. E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4, 1–30.
- Bates, D. S., 2000. Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics* 94, 181–238.
- Becker, K. G., Finnerty, J. E., Gupta, M., 1990. The intertemporal relation between the u.s. and japanese stock markets. *The Journal of Finance* 45, 1297–1306.
- Billio, M., Getmansky, M., Lo, A. W., Pelizzon, L., 2010. Measuring systemic risk in the finance and insurance sectors. Tech. rep., 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.
- Carr, P., Wu, L., 2003. What type of process underlies options? A simple robust test. *The Journal of Finance* 58, 2581–2610.
- Corradi, V., Distaso, W., Fernandes, M., 2012. International market links and volatility transmission. *Journal of Econometrics*, forthcoming .
- Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159, 276–288.
- Crane, R., Sornette, D., 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., Fry, R., Gonzalez-Hermosillo, B., Martin, V. L., 2010. Sampling properties of contagion tests. Tech. rep., University of Melbourne.
- Dungey, M., Gonzalez-Hermosillo, B., 2005. Empirical modeling of contagion: A review of methodologies. *Quantitative Finance* 5, 9–24.
- Ehrmann, M., Fratzscher, M., Rigobon, R., 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, 1367–1404.
- Errais, E., Giesecke, K., Goldberg, L. R., 2010. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics* 1, 642–665.
- Fan, Y., Fan, J., 2011. Testing and detecting jumps based on a discretely observed process. *Journal of Econometrics* 164, 331–344.
- Hamao, Y., Masulis, R. W., Ng, V., 1990. Correlations in price changes and volatility across international stock markets. *Review of Financial Studies* 3, 281–307.
- 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, B* 33, 438–443.
- Hawkes, A. G., 1971b. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 83–90.
- Hawkes, A. G., Oakes, D., 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.
- Huang, X., Tauchen, G. T., 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 4, 456–499.
- Jiang, G. J., Oomen, R. C., 2008. Testing for jumps when asset prices are observed with noise – A “swap variance” approach. *Journal of Econometrics* 144, 352–370.
- 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. S., Hannig, J., 2010. Detecting jumps from Lévy jump diffusion processes. *Journal of Financial Economics* 96, 271–290.
- Lee, S. S., Mykland, P. A., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* 21, 2535–2563.
- Maheu, J. M., McCurdy, T. H., 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., Short, M. B., Brantingham, P. J., Schoenberg, F. P., Tita, G. E., 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., Akaike, H., 1982. On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes. *Journal of the Royal Statistical Society, 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., Schbath, S., 2010. Adaptive estimation for Hawkes processes; Application to genome analysis. *Annals of Statistics* 38, 2781–2822.
- Salmon, M., Tham, W. W., 2007. Time deformation and the yield curve. Tech. rep., 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.

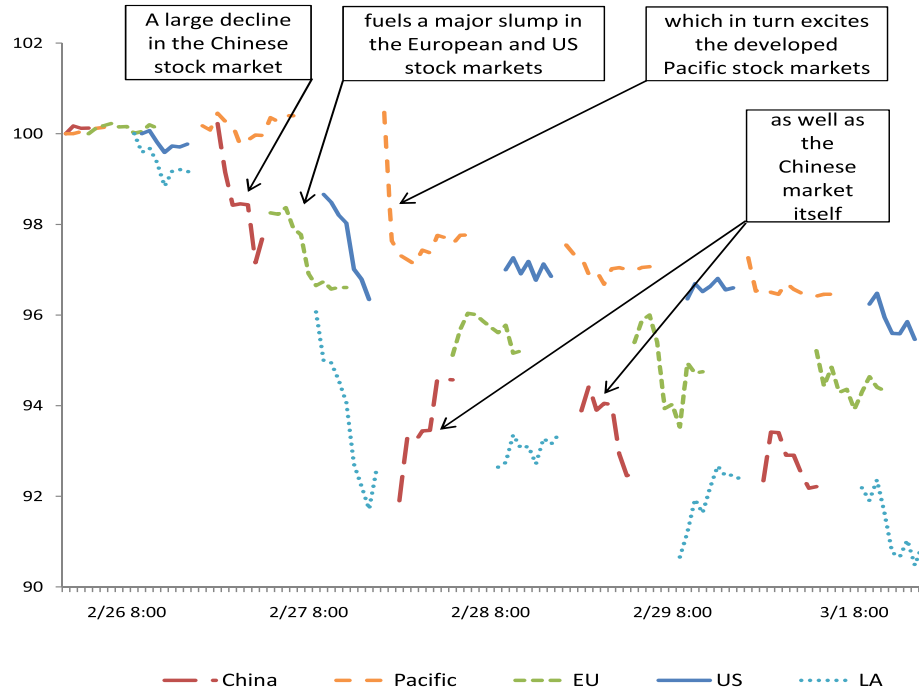


Fig. 1. Mutual Excitation: Example I.

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.

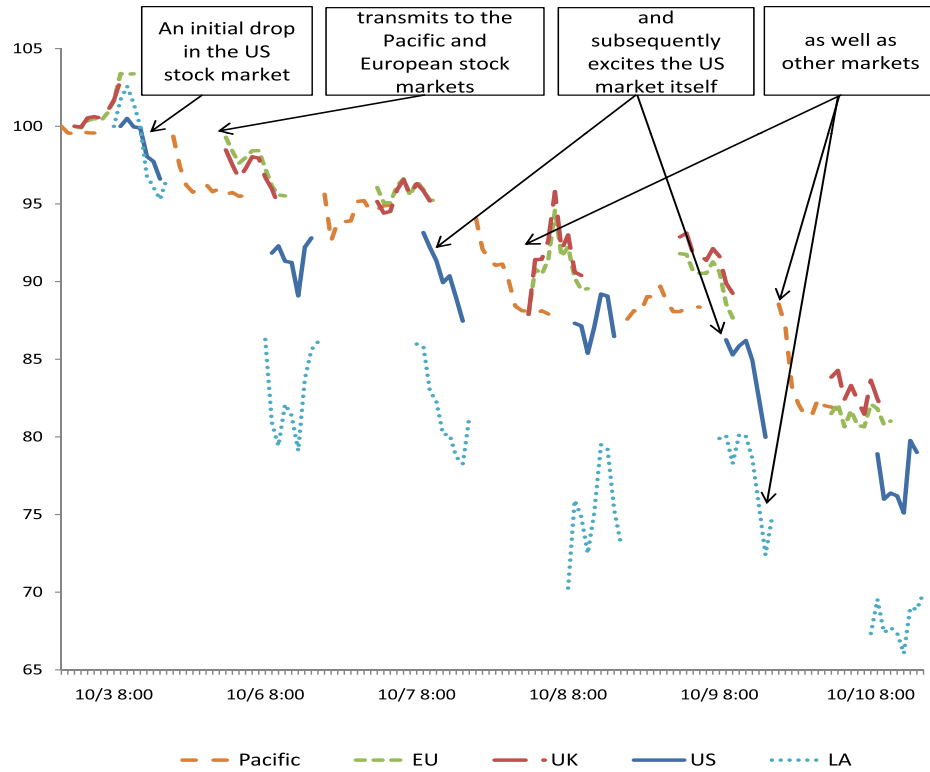


Fig. 2. Mutual Excitation: Example II.

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.

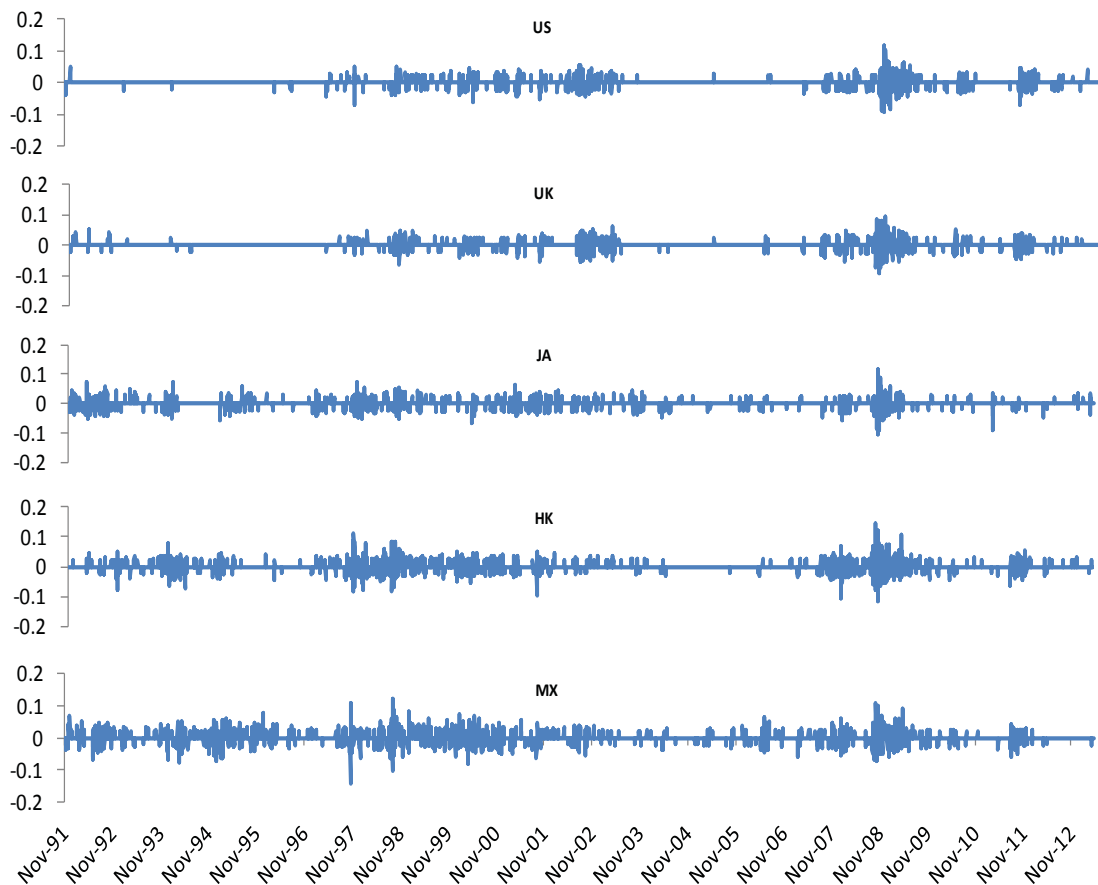


Fig. 3. Large Stock Index Returns in the Five World Regions.

This figure plots the large log-returns of open-to-close and close-to-open international equity index data. We consider five series: S&P 500 (US); FTSE 100 (UK); Nikkei 225 (Japan, JA); Hang Seng (Hong Kong, HK); and IPC (Mexico, MX). The (joint) sample period considered in this figure is November 8, 1991 to April 30, 2013. 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 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 and the full individual sample periods.

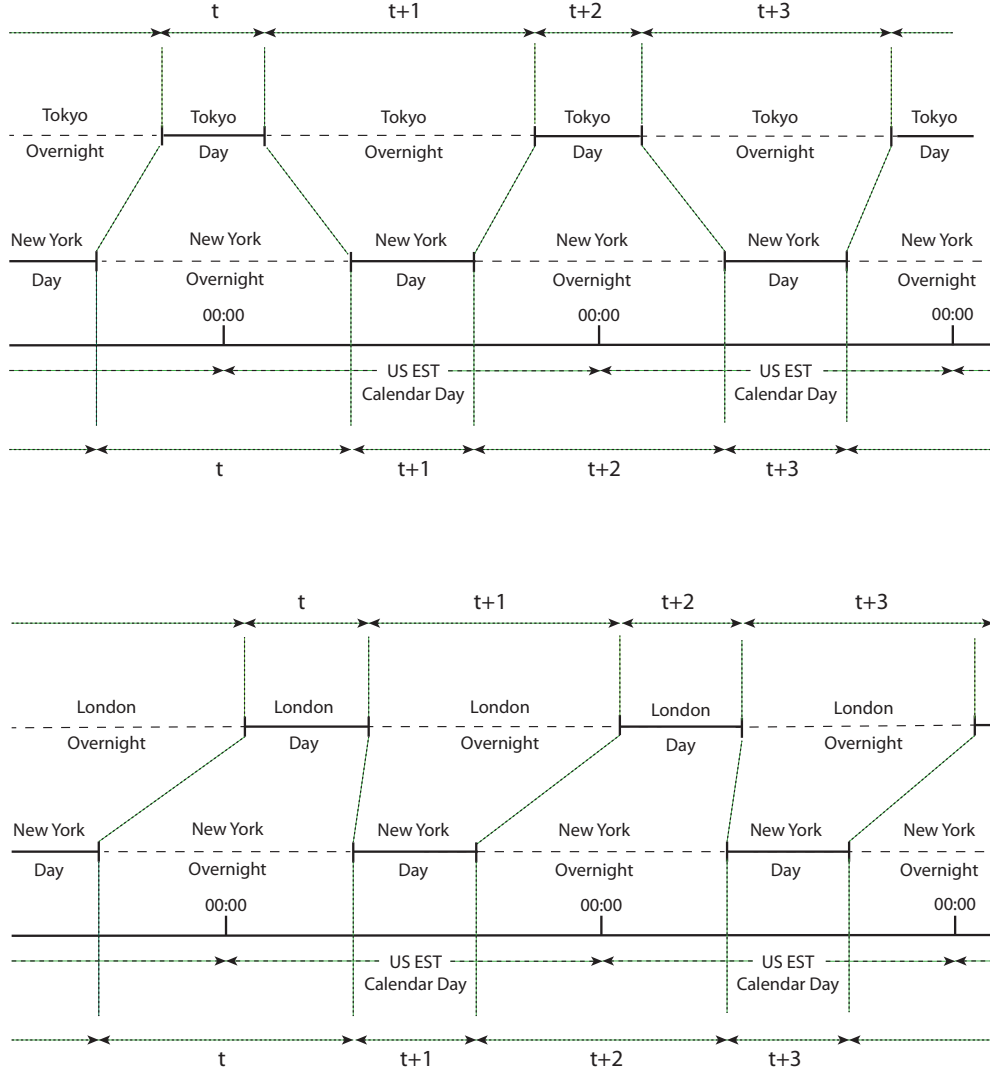


Fig. 4. International Data Synchronization.

This figure visualizes our treatment of the open and close international equity index data. For the US and Japan, we define as instantaneous the overnight return in Japan and the corresponding day return in the US; and the day return in Japan and the corresponding overnight return in the US. Similarly, for the US and UK, we define as instantaneous the overnight return in the UK and the corresponding day return in the US; and the day return in the UK and the corresponding overnight return in the US.

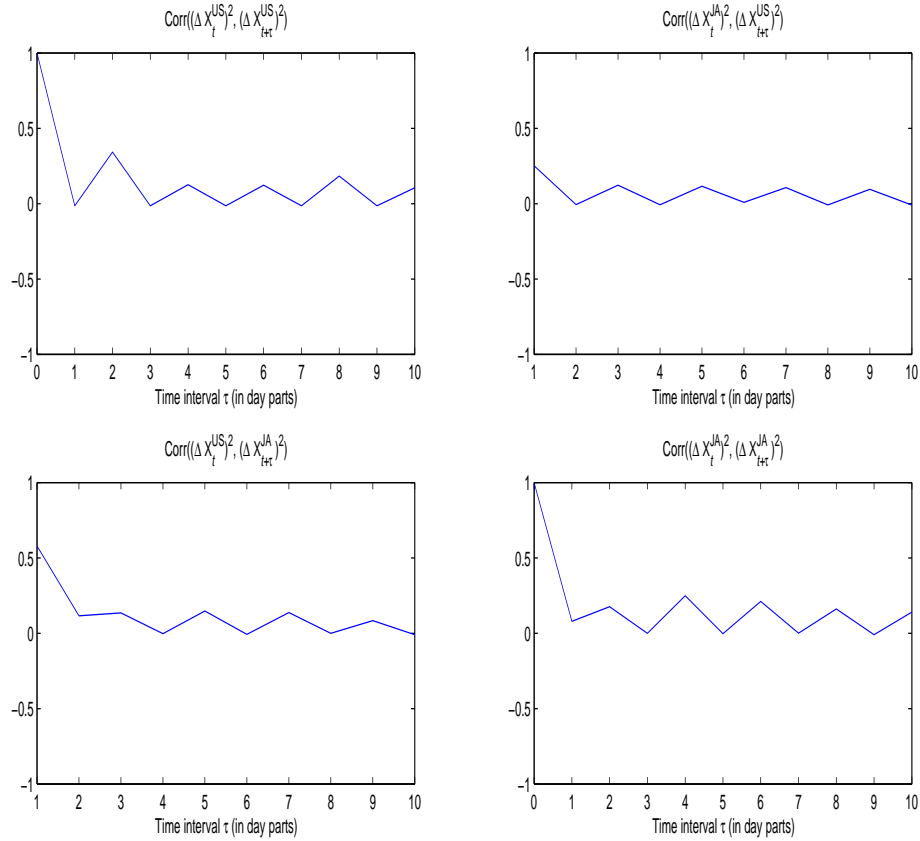


Fig. 5. Autocovariances and Cross-Covariances of Squared Returns US and Japan.

This figure plots autocovariances and cross-covariances of the synchronized squared log-returns of the open and close international equity data for the pair US and Japan. The unit of the index  $\tau$  is number of day parts (half days). As expected, autocorrelations between day-and-day squared returns (corresponding to lags 0, 2, 4, 6, ...) are larger than autocorrelations between day-and-night squared returns (corresponding to lags 1, 3, 5, 7, ...). Similarly, cross-correlations between day-and-day squared returns (corresponding to lags 1, 3, 5, 7, ...) are larger than cross-correlations between day-and-night squared returns (corresponding to lags 2, 4, 6, 8, ...). This manifests itself in the “seesaw” pattern observed in the plot. It is also apparent that autocorrelations and cross-correlations of squared returns die out progressively as predicted by the exponential decay implied by the model.



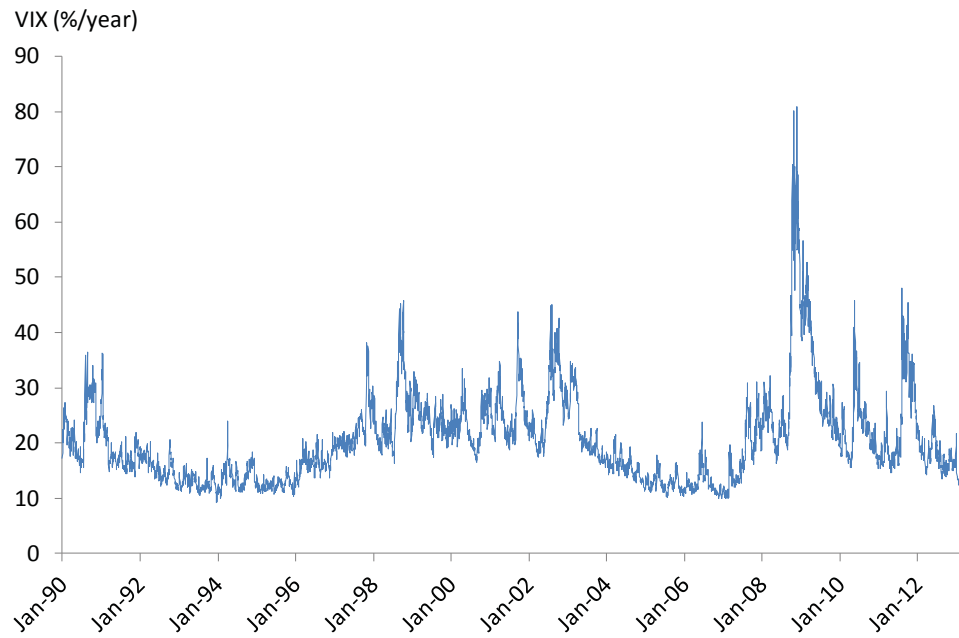


Fig. 6. Volatility Index (VIX).

This figure plots the close index values of the VIX index, viewed as a measure of total risk including continuous and jump risk. Sample period: January 2, 1990 to April 30, 2013.

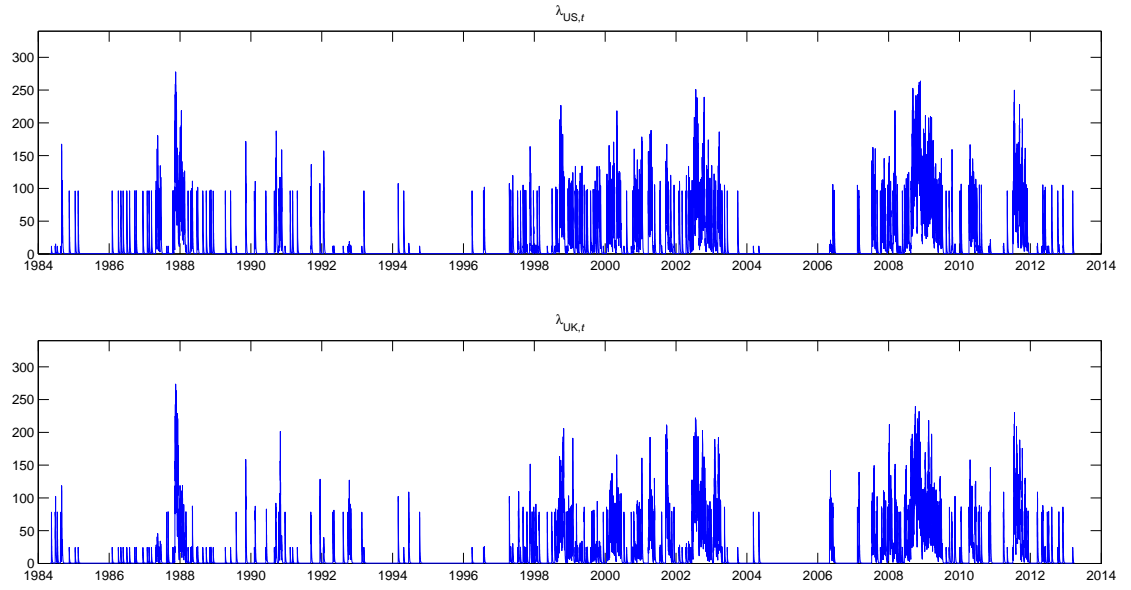


Fig. 7. Estimated Time Series of the US and UK Jump Intensities.

This figure plots the estimated time series of  $\lambda_{US,t}$  and  $\lambda_{UK,t}$ , filtered from the time series of asset returns. The adopted filtering procedure is described in Section 4.5. Sample period: April 2, 1984 to April 30, 2013.

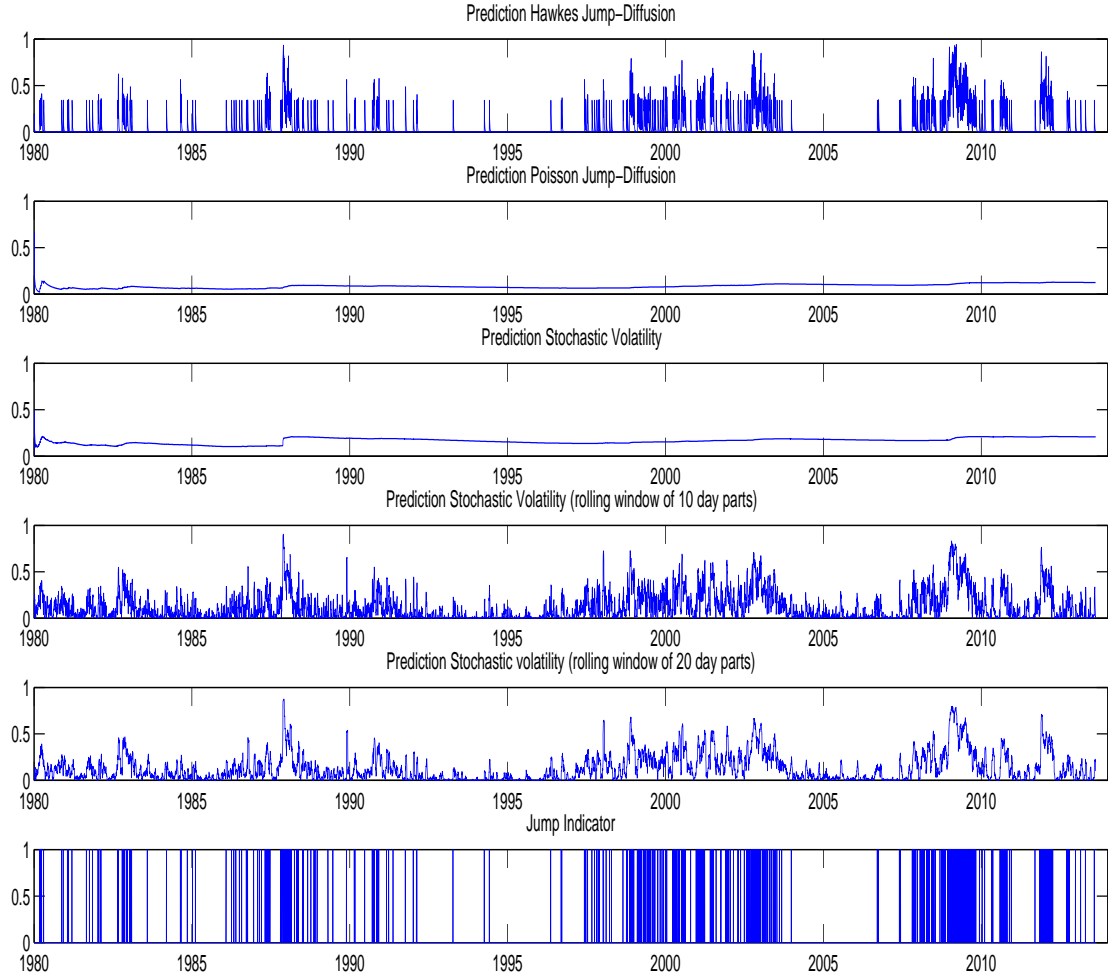


Fig. 8. Out-of-Sample Prediction of Jump Probabilities.

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  $\Delta = 1$  full 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 larger than 2% in absolute value, given a realized volatility estimate (based on the full history or on a rolling window). The jump indicator takes the value one if the US index return is larger in absolute value than 2%, and zero otherwise. The respective root Mean Squared Prediction Errors are 31.14 (Hawkes Jump-Diffusion), 31.34 (Poisson Jump-Diffusion), 33.94 (Stochastic Volatility), 32.63 (Stochastic Volatility — 10-day parts rolling window), 32.11 (Stochastic Volatility — 20-day parts rolling window).

	US80	US90	UK	JA	HK	MX
Mean	0.00016	0.00013	0.00012	0.00002	0.00017	0.00032
St. Deviation	0.0080	0.0081	0.0080	0.0095	0.0117	0.0111
Skewness	-1.69	-0.30	-0.60	-0.34	-2.11	0.06
Excess Kurtosis	57.1	18.4	20.8	18.6	53.2	14.1

Table 1. Descriptive Statistics: International Equity Indices.

This table reports descriptive statistics for the log-returns of open and close international equity index data collected from finance.yahoo.com. We consider five indices: S&P 500 (US); FTSE 100 (UK); Nikkei 225 (Japan, JA); Hang Seng (Hong Kong, HK); and IPC (Mexico, MX). Sample period considered: January 2, 1980 (for US80), January 2, 1990 (for US90), April 2, 1984 (for UK), January 4, 1984 (for JA), December 31, 1986 (for HK), and November 8, 1991 (for MX). Our sample period ends on April 30, 2013, for all indices.

	US80	US90	UK	JA	HK	MX
$\alpha$	105.8*** (20.0)	85.5** (37.5)	109.0*** (23.9)	123.2** (52.3)	99.5*** (17.6)	92.3*** (26.1)
$\beta$	94.1*** (14.0)	81.6** (35.6)	101.0*** (23.0)	111.2** (51.5)	81.4*** (16.6)	85.6*** (23.8)
$\lambda$	6.31*** (0.10)	2.39*** (0.01)	2.22*** (0.01)	4.63*** (0.05)	23.93*** (0.76)	5.13*** (0.02)
$\lambda_\infty$	0.70	0.11	0.16	0.45	4.36	0.37
$\sqrt{\theta}$	0.141*** (0.0276)	0.169*** (0.0099)	0.166*** (0.0082)	0.193*** (0.0164)	0.145 (0.0939)	0.226*** (0.0092)
$\mu$	0.161*** (0.0623)	0.080*** (0.0290)	0.087*** (0.0271)	0.039 (0.0402)	0.380** (0.1854)	0.153*** (0.0582)
$1/\gamma$	0.030	0.030	0.031	0.030	0.031	0.032
$p$	0.711	0.605	0.694	0.599	0.699	0.480

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

This table reports the GMM parameter estimates for the parameters of the univariate 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: US80; US90; UK; Japan (JA); Hong Kong (HK); and Mexico (MX). Sample period considered: Starting January 2, 1980 (for US80), January 2, 1990 (for US90), April 2, 1984 (for UK), January 4, 1984 (for JA), December 31, 1986 (for HK), and November 8, 1991 (for MX), and ending on April 30, 2013, for all indices.

1	US	US	US	US
2	UK	JA	HK	MX
$\alpha$	115.0*** (17.5)	116.8*** (30.9)	98.4*** (22.1)	118.9*** (45.4)
$\beta_{11}$	95.5*** (17.4)	98.0*** (28.7)	78.7*** (20.1)	114.1*** (41.8)
$\beta_{12}$	11.2 (28.4)	10.0 (8.7)	0.1 (2.7)	0.0 (2.8)
$\beta_{21}$	23.8** (9.6)	14.6** (6.4)	20.0** (9.2)	9.9*** (2.5)
$\beta_{22}$	77.7*** (12.8)	92.6*** (34.8)	78.4*** (21.3)	108.3** (43.3)
$\lambda_1$	6.11*** (0.045)	6.73*** (0.054)	10.90*** (0.377)	1.71*** (0.002)
$\lambda_2$	5.46*** (0.050)	6.58*** (0.055)	21.49*** (0.191)	2.36*** (0.002)
$\lambda_\infty$	0.50	0.52	2.15	0.07
$\sqrt{\theta_1}$	0.146*** (0.0154)	0.150*** (0.0308)	0.139*** (0.0238)	0.179*** (0.0081)
$\sqrt{\theta_2}$	0.149*** (0.0201)	0.188*** (0.0239)	0.162** (0.0775)	0.242*** (0.0085)
$\rho$	0.219** (0.1090)	0.357*** (0.1259)	0.894 (0.6150)	0.603*** (0.0503)
$\mu_1$	0.164*** (0.0451)	0.141*** (0.0509)	0.206*** (0.0603)	0.078** (0.0333)
$\mu_2$	0.126*** (0.0426)	0.052 (0.0487)	0.353** (0.1710)	0.161*** (0.0610)
$1/\gamma_1$	0.031	0.031	0.030	0.030
$1/\gamma_2$	0.031	0.030	0.031	0.032
$p_1$	0.721	0.643	0.700	0.607
$p_2$	0.689	0.600	0.701	0.481

Table 3. Parameter Estimates for the Bivariate Hawkes Jump-Diffusion Model.

This table reports the GMM parameter estimates for the 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 four series in pairs with the US: UK; Japan (JA); Hong Kong (HK); and Mexico (MX). Sample period considered: Starting April 2, 1984 (for UK), January 4, 1984 (for JA), December 31, 1986 (for HK), and November 8, 1991 (for MX), and ending on April 30, 2013, for all indices.

	Null Hypothesis	US UK	US JA	US HK	US MX
$\mathcal{H}_0^I$	No Excitation	***	***	***	***
$\mathcal{H}_0^{II}$	No Self-Excitation	***	***	***	***
$\mathcal{H}_0^{III}$	No Cross-Excitation	***	**	**	***

Table 4. 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 five series: US; UK; Japan (JA); Hong Kong (HK); and Mexico (MX).