THE HAWKES PROCESS AND TIME-VARYING JUMP INTENSITY IN FINANCIAL TIME SERIES

Maciej Kostrzewski

Abstract

News might trigger arrivals of jumps in financial time series. The Bayesian JD(M)J model is applied to detect jumps. The Bayesian framework, founded upon the idea of latent variables and computationally facilitated with Markov Chain Monte Carlo methods, enables the detection of jumps and the analysis of their frequency.

The presented methodology is illustrated with empirical studies employing both simulated and real-world datasets. A very intuitive observation is made, namely that higher posterior probabilities of jumps are inferred during the periods of higher absolute values of returns. A series of waiting times between two consecutive jumps is also of interest in the study. Periods of no jumps alternating with the ones of frequent jumps confirm the existence of jump clustering.

The above results may prompt one to apply Hawkes processes to model the moments when jumps occur. The results of the maximum likelihood estimation of Hawkes process, again, indicate the jump clustering phenomenon. Information criteria point to a major superiority of the models featuring a stochastic intensity of jumps.

Key words: jump clustering, Bernoulli jump-diffusion model, Hawkes processes, Bayesian inference, maximum likelihood estimation

JEL Code: G17, C58

Introduction

Numerous studies indicate that many financial time series feature drastic occasional movements (referred to as jumps), although, obviously, whether to classify a given observation as a one featuring a jump or not typically hinges on some arbitrary definition of a jump itself.

One of the most common group of models employed in modeling time series that consist of "typical" (and continuous) changes and, simultaneously, allow for abnormal occasional shifts (jumps) is the class of jump-diffusion processes. In the current research, a very specific instance of these is under consideration, namely the Jump-Diffusion with *M* Jumps (henceforth JD(*M*)J) process, developed by (Kostrzewski, 2012a, 2012b, 2013a), and derived by discretizing some jump-diffusion process. To estimate the parameters of the JD(*M*)J model and to infer about jumps, we resort to Bayesian statistical framework, equipped with the MCMC methods

The very term "jump clustering" – quite analogous to the one of "volatility clustering", pervading the GARCH and SV literature (see, e.g. (Osiewalski, Pipień, 2004), (Pajor, 2009))– means that jump arrivals (or waiting times between two consecutive jumps) tend to cluster, i.e. if a given jump arrives in a short time since the previous one, then, most possibly, another jump will follow soon, too. Jump clusters have already been discussed in the financial econometrics literature – see, e.g., (Knight and Satchell, 1998), (Maheu and McCurdy, 2004), (Lee, 2012). The main idea is based on the assumption of a stochastic jump intensity which follows, e.g., a self-exciting process.

In what follows, a very common (though still arbitrary) rule of classifying a given observation as a jump is adhered to, according to which a data point is diagnosed as a jump if the posterior probability of a jump exceeds 0.5. Then, a series of waiting times between two consecutive jumps is formed, enabling one to examine the jump clustering phenomenon. Furthermore, the series of time moments of jumps is fit (via a maximum likelihood approach) with a Hawkes process, which provides a method of analyzing the time-variability of jump intensity. The latter also facilitates the detection of jump clusters. Such an (eclectic, indeed) approach is employed in this study to identify jumps and their clusters in the case of both simulated and real-world datasets.

For relevant definitions and properties of the Hawkes processes we refer the reader to (Hawkes, 1971a, 1971b). The non-Bayesian estimation of the models founded upon these processes is discussed by (Ogata, 1978), (Daley and Vere-Jones, 2003), whereas the Bayesian approach is exposed by (Rasmussen, 2013). The Hawkes processes are applied to many fields including seismology, sociology, neuroscience and others, also including financial time series analysis (see (Ait-Sahalia, Cacho-Diaz and Laeven, 2013); (Maneesoonthorn, Forbes and Martin, 2014)).

The contribution of the paper resides in performing an analysis of time-varying jump frequency and designing a visual method of detecting the jump clustering phenomenon. The use of the proposed methodology of detecting jumps, for some time series under consideration, is particularly justified in the context of settling whether the jump clustering phenomenon should or need not be explicitly incorporated into the structure of some common jump-diffusion models.

1 The JD(M)J model

Consider a standard Wiener process $W = (W_t)_{t \ge 0}$, a Poisson process $N = (N_t)_{t \ge 0}$ with a constant intensity parameter $\lambda > 0$, and the sequence of independent random variables $Q = (Q_j)_{j \ge 1}$. Let us assume that W, N and Q are mutually independent. Finally, $S = (S_t)_{t \ge 0}$ denotes the price process of some risky asset.

The logarithm of S_t is governed by a jump-diffusion process that constitutes the solution of the equation:

$$d(\ln S_t) = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t + QdN_t.$$

It might be shown that:

$$\ln\left(\frac{S_{t+\Delta}}{S_t}\right) = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta + \sigma(W_{t+\Delta} - W_t) + \sum_{i=N_t+1}^{N_{t+\Delta}} Q_i, \quad \Delta > 0.$$

The process is built of two components: the (pure) diffusion part $\left(\mu - \frac{1}{2}\sigma^2\right)\Delta + \sigma(W_{t+\Delta} - W_t)$, representing continuous variations in the series, and the (pure) jump component, $\sum_{i=N_t+1}^{N_{t+\Delta}} Q_i$, reflecting abnormal (extreme) movements in returns. The continuous price behavior between jumps is described by the geometric Brownian motion, W, while the arrival rate of jumps is described by the homogeneous Poisson process, N, and the jump magnitudes – by Q.

The distribution of logarithmic rates of return, $\ln\left(\frac{S_{t+\Delta}}{S_t}\right)$, is an infinite mixture:

$$\sum_{k=0}^{\infty} \exp(-\lambda\Delta)\frac{(\lambda\Delta)^k}{k!} f_k(x), \tag{1}$$

where f_k are some densities. Since the series given by (1) is infinite, the density is intractable. Therefore, consider the following finite approximation of (1):

$$\sum_{k=0}^{\infty} \exp(-\lambda\Delta)\frac{(\lambda\Delta)^k}{k!} f_k \approx \sum_{k=0}^{M} \exp(-\lambda\Delta)\frac{(\lambda\Delta)^k}{k!} f_k. \tag{2}$$

The approximation restricts the number of jumps over any time interval Δ to M. The case of

$M=0$ indicates no jumps over interval Δ. To obtain the conditional data density (given the vector of parameters θ) the approximation is normalized.

Further considerations are restricted to the discrete time framework. Time series $(x_1, x_2, \ldots, x_n)$ is comprised of $x_i = \ln\left(\frac{S_{t_{i+1}}}{S_{t_i}}\right)$ observed at times $t_1, t_2, \ldots$. Moreover, $\Delta \equiv t_{i+1} - t_i > 0$ is a fixed time interval between following observations. The specification, termed the JD(2)J model, is defined by assuming a normal distribution for Q_j: $f_{Q_j}(x) = \frac{1}{\sqrt{2\pi\sigma_Q}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_Q}{\sigma_Q}\right)^2\right)$, and setting $M=2$. The model is used to model series of daily logarithmic rates of return of S&P100 index, indicating $\Delta = \frac{1}{252}$. The choice of $M=2$ is explained in (Kostrzewski, 2013a).

2 The Bayesian JD(2)J model

A Bayesian statistical model is defined by the joint density: $p(x,\theta) = p(x|\theta)p(\theta)$, where $x = (x_1, \ldots, x_n)$ is the observed data, θ is the vector of unknown parameters, $p(x|\theta)$ is the sampling density and $p(\theta)$ is the prior density. The inference rests upon the posterior density $p(\theta|x)$ of θ given data x.

Under the JD(2)J specification the process depends on five unknown parameters $\theta = (\mu, \sigma, \lambda, \mu_Q, \sigma_Q^2)$, where $\theta \in \mathsf{R} \times (0,\infty) \times (0,\infty) \times (0,1) \times (0,\infty) \times (0,\infty)$.

When we analyze time series which is believed to be a trajectory of some JD(*2*)J process then one does not really know if a given data point observation has been generated by the pure diffusion or the jump-diffusion component. In other words, one cannot determine which component of the series in (2), i.e. $\phi\left(\cdot; \left(\mu - \frac{1}{2}\sigma^2\right)\Delta + \mu_Q k, \sigma^2\Delta + \sigma_Q^2 k\right), k \in \{0,1,2\}$ is responsible for the observation. To manage the problem we introduce latent variables $Z = (Z_1, \ldots, Z_n)$ such that $Z_i \in \{0,1,2\}$ and $P(Z_i = j) = w_j$ where $i \in \{1, \ldots, n\}$ and $j \in \{0,1,2\}$. The value $Z_i = 0$ means no jump at (an interval) $t = i\Delta$. The values $Z_i = 1$ and $Z_i = 2$ imply that a jump occurs at (an interval) $t = i\Delta$ and its value distribution is $\phi(\cdot; \mu_Q, \sigma_Q^2)$ and $\phi(\cdot; 2\mu_Q, 2\sigma_Q^2)$, respectively. In other words, the events $Z_i = 1$ and $Z_i = 2$ correspond to a "smaller" and "larger" jump (in terms of absolute value), respectively. By means of Z one can detect times of jumps. Now, the Bayesian JD(M)J model – enhanced with latent variables Z – is given by:

$$p(x,\theta,Z) = p(x|\theta,Z)\,p(\theta,Z).$$

Formally, the occurrence of a jump is equivalent to the event $Z_i > 0$. Unfortunately, one does not observe Z_i, but the posterior probability of a jump, $P(Z_i > 0|x)$, can be evaluated for each $i = 1,...,n$. Let us assume that a jump occurs at the *i*-th period if the probability $P(Z_i > 0|x)$ exceeds an arbitrarily chosen value of 0.5. The resulting series consisting of zeros and ones corresponding to such *i*'s that $P(Z_i = 0|x) \leq 0.5$ and $P(Z_i > 0|x) > 0.5$, respectively, is further employed in studying the jump clustering phenomenon.

Posterior characteristics of all unknown quantities are calculated via the Markov Chain Monte Carlo (MCMC) methods, combining the Gibbs sampler, the independence and the sequential Metropolis-Hastings algorithms, as well as the acceptance-rejection sampling (Gamerman and Lopes, 2006). The details on the prior structure introduced into the model, the adopted MCMC methods and the results of estimation could be find in (Kostrzewski, 2013a).

3 The one-dimensional Hawkes model

Let us consider the simple point process *N* in time domain. The simple point process N might be specified by conditional intensity $\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E(N(t,t+\Delta t) \mid \Im_t)$, where $\Im_t$ represents the history of the process *N* up to *t* and $N(t,t+\Delta t)$ is the number of points (jumps) in the interval $(t,t+\Delta t)$. The conditional intensity $\lambda(t)$ is interpreted as the instantaneous rate of occurrence of events at time *t*. A well-known example of the simple point process is the (point) Poisson process. The homogeneous Poisson process counts events that occur at a constant rate λ, whereas the non-homogeneous one counts events that occur at a variable (time-dependent) rate $\lambda(t)$. An expected number of events (jumps) over a finite interval $(0,T)$ is $E(N(0,T)) = \int_0^T \lambda(t)dt$.

The Hawkes process is the simple point process. The one-dimensional Hawkes-type cluster model (Daley and Vere-Jones, 2003) for the times of events (jumps) considered in the paper is an example of the classical linear Hawkes process specified by the conditional intensity :

$$\lambda(t) = \lambda_0 + \int_{-\infty}^{t} g(t-s)N(ds) = \lambda_0 + \sum_{t_i < t} g(t-t_i), \tag{3}$$

where $g(z)=\sum_{k=1}^{K} a_k z^{k-1} e^{-cz}$ is usually referred to as the exciting function. Moreover $\lambda_0 > 0$, $a_k > 0$, $c > 0$. These inequalities ensure that the conditional intensity is positive. The parameter λ_0 represents the background rate of occurrence, i.e. the intensity if there have been no past events. The parameters a_k and *c* control the level of clustering. *K* represents the order of the exciting function. Note that $\lambda(t)$ is a stochastic process. The intensity of the process depends on the entire history and is self-exciting. The process has the clustering property which is a consequence of the self-excitation feature. Such processes are broadly employed in the literature on, e.g., the occurrence of earthquakes.

It has been shown that under general conditions, the maximum likelihood estimates of simple point processes are consistent and asymptotically normal (see: (Ogata, 1978), (Daley and Vere-Jones, 2003)).

4 Examples

In this section we illustrate the methodology presented above. Two datasets are under consideration: a simulated and a real-world one. In both cases we perform Bayesian estimation of the model in question by means of the author's own algorithms programmed in R (R Core Team, 2013), whereas the maximum likelihood estimates of the Hawkes model's parameters are obtained via the R package ptproc (Peng, 2002).

4.1 Simulation case study

The series of n = 100 random data points generated from the uniform distribution over an interval [0,3] is under consideration. Simulated values are perceived as time moments of jumps occurring over three years. Note that this implies an average of thirty jumps per year. According to the way the data are generated, the jumps do not manifest themselves in clusters. Below we examine the results of fitting (via the maximum likelihood estimator) the simulated series with the Hawkes process under $K = 1$ (see Table 1). A close-to-zero value of the estimate of α indicates no jump clustering. What is more, $\hat{\lambda}_0 \approx 33.9$, which is close to the actual expected number of jumps per year.

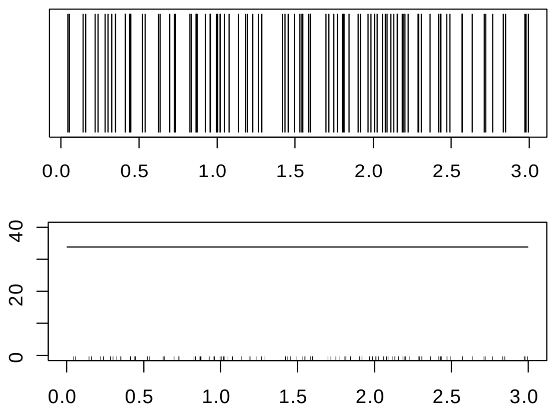
Tab. 1: The MLE estimates of the Hawkes process for *K* = 1 and simulation data

θ	λ_0	a	c
$\hat{\theta}_{MLE}$	33.9	2.505e-08	5.534

Source: own elaboration.

Introducing the parameters' estimates into (3) yields the expression of jumps' conditional intensity: $\lambda(t) = 33.9 + \sum_{t_i<t} 2.505*10^{-8} e^{-5.534(t-t_i)}$. Since the summation term in this formula is close to zero, the conditional intensity of jumps is almost constant: $\lambda(t) \approx \lambda_0 = 33.9$. We also estimated a model for $K = 0$ (i.e. the homogeneous Poisson process). According to the values of the Akaike information criterion calculated for both specification: –498.7524 for $K = 1$, and –502.7524 for $K = 0$, the data supports the simpler model structure. Figure 1 presents the time moments of jumps (top) and the corresponding values of $\lambda(t)$ with indicated moments of jumps (bottom).

Fig. 1: The time moments of jumps (top) and the values of the conditional intensity (with indicated moments of jumps; bottom) under *K* = 1 (simulated data)



Source: own elaboration.

4.2 S&P100 Index

To illustrate the methodology presented in the paper, we also analyze a series of daily logarithmic rates of return on the S&P100 Index over the period from March 5, 1984 through July 8, 1997. The series has already been employed by (Honore, 1998), who fits it with the Bernoulli jump-diffusion model by means of the maximum likelihood method, as well as by (Kostrzewski, 2013a). Quotations on the S&P100 Index have been downloaded from (EconStats, 2012).

In Table 2 values of the Akaike information criterion are presented for models featuring different values of $K \in \{0, 1, \ldots, 4\}$. The results support the Hawkes process with $K = 1$. Table 3 displays the MLE estimates of the Hawkes process' parameters (under $K = 1$) along with their standard errors.

Tab. 2: Akaike information criterion values for various *K* and the series of daily logarithmic rates of return on the S&P100 Index

K	0	1	2	3	4
AIC	-106.49	-200.33	-140.27	-8.57	91.91

Source: own elaboration.

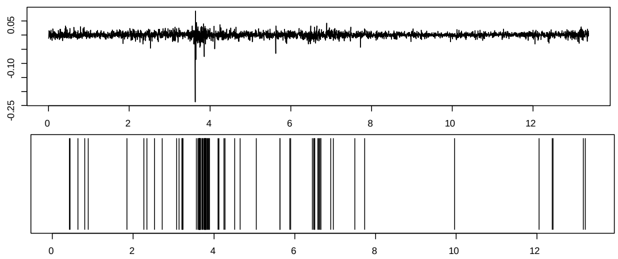
Tab. 3: The MLE estimates and standard errors of the Hawkes process' parameters under *K* = 1 and the returns on the S&P100 Index

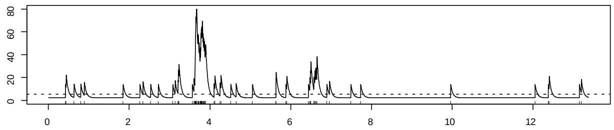
θ	λ_0	a	c
Estimates (MLE)	2.283	12.61	20.601
Standard errors	0.536	0.491	8.56

Source: own elaboration.

Introducing the parameters' estimates into (3) yields the expression of jumps' conditional intensity: $\lambda(t) = 2.283 + \sum_{t_i < t} 12.61 e^{-20.601(t-t_i)}$. The estimate $\hat{\lambda}_0 = 2.283$ represents the background rate of occurrence, i.e. the intensity if there have been no past events. The assessed value of a: $\hat{a} = 12.61$ indicates that immediately after a jump, the conditional intensity increases by about 12.61 events per day and implies jump clustering. Note that in longer periods without jumps the conditional intensity $\lambda(t) \approx \hat{\lambda}_0 = 2.283$. The MLE estimate of the constant intensity under the homogeneous Poisson process ($\hat{\lambda} = 5.45$) is twice as large as $\hat{\lambda}_0$. A larger value of $\hat{\lambda}$ may result from the occurrence of periods featuring higher intensity of jumps, these being not incorporated in the homogeneous Poisson process.

Fig. 2: Logarithmic rates of return on the S&P100 Index (top), the time moments of jumps (middle) and the values of the conditional intensity (with indicated moments of jumps and a dotted line representing the value of $\hat{\lambda} = 5.45$; bottom) under *K* = 1





Source: own elaboration.

Figure 2 plots the modeled series of logarithmic returns on the S&P100 Index, the time moments of jumps (middle) and the values of the conditional intensity (with indicated moments of jumps and a dotted line representing the value of $\hat{\lambda} = 5.45$; bottom) under $K = 1$. One can easily observe periods of no jumps alternating with the ones of frequent jumps, clearly indicating the time-variability of jump's intensity and the jump clustering phenomenon.

Final remarks

Two major conclusions can be drawn from the research. Firstly, the jumps may manifest themselves in clusters (the jump clustering phenomenon). Secondly, as it follows from the latter it may be empirically more justifiable to allow for the stochasticity of the jump intensity, instead of restricting the intensity to be constant throughout.

It is worth noting that even though the JD(2)J model – similarly as some other common specifications – does not account for any dependence structure in the occurrence of jumps, it is still informative (in the context of detecting jump clusters) to inspect the series of time moments when jump occur and the series of times elapsed between consecutive jumps, for it still can exhibit patterns suggestive of clustering. Specifically, via the approach proposed in the study it has been shown that the series of returns on the S&P100 exhibit clusters of jumps. The jump frequency analyses of some other time series, not mentioned here, also support the time-varying intensity of jumps and confirm the existence of jump clustering. Naturally, such results incline one to relax the assumption of a constant jump intensity and to allow for some (stochastic) time-variability of this parameter.

Acknowledgements

Support by the Faculty of Management, Cracow University of Economics is acknowledged. Useful comments and remarks by anonymous referee are highly appreciated. The author would also like to thank Łukasz Kwiatkowski for the language verification of the manuscript.

References

Ait-Sahalia, Y., Cacho-Diaz, J., & Laeven, R.J.A. (2013). *Modeling Financial Contagion Using Mutually Exciting Jump Processes*. Working Paper. Retrieved April 21, 2014, from http://www.princeton.edu/~yacine/contagion.pdf

Daley, D. J., & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods* (2 ed.). New York: Springer.

EconStats: GDP Industrial Production Money Supply Inflation US Japan China. Retrieved 2012, from http://www.econstats.com.

Gamerman, D., & Lopes H.F. (2006). Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference. Boca Raton: Chapman & Hall/CRC. 2006.

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

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

Honore, P. (1998, January 24). *Pitfalls in Estimating Jump-Diffusion Models*. Retrieved April 8, 2014, from http://dx.doi.org/10.2139/ssrn.61998.

Knight, J.L., & Satchell, S. (2007). GARCH Processes – Some Exact Results, Some Difficulties and a Suggested Remedy. [In:] Knight, J. L., & Satchell, S. (eds.) (2007). *Forecasting Volatility in the Financial Markets*. Amsterdam: Elsevier.

Kostrzewski, M. (2012a). Bayesian Pricing of the Optimal-Replication Strategy for European Option in the JD(M)J Model. *Dynamic Econometric Models*, *vol. 12*, 53-71.

Kostrzewski, M. (2012b). On the Existence of Jumps in Financial Time Series. *Acta Physica Polonica B*, *vol. 43*, 2001-2019.

Kostrzewski, M. (2013a). Bayesian Inference for the Jump-Diffusion Model with M Jumps. *Communications in Statistics – Theory and Methods*. DOI: 10.1080/03610926. 2012.755202.

Lee, S.S. (2012). Jumps and Information Flow in Financial Markets. *Review of Financial Studies*, *vol. 25*, 439-479.

Maheu, J.M., & McCurdy, T.H. (2004). News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns. *The Journal of Finance*, *vol. 59*, 755-793.

Maneesoonthorn, W., Forbes, C., & Martin, G. (2014, January 16). *Inference on Self-Exciting Jumps in Prices and Volatility using High Frequency Measures*. Working Paper. Retrieved April 21, 2014, from http://arxiv.org/pdf/1401.3911v1.pdf

Ogata, Y. (1978). The Asymptotic Behaviour of Maximum Likelihood Estimators for Stationary Point Processes. *Annals of the Institute of Statistical Mathematics*, *30*, 243-261.

Osiewalski J., & Pipień M. (2004). Bayesian Comparison of Bivariate ARCH-TYPE Models for the Main Exchange Rates in Poland, *Journal of Econometrics, vol. 123*, 371-391

Pajor A. (2009). Bayesian analysis of the Box-Cox transformation in Stochastic Volatility models. *Dynamic Econometric Models*, vol. 9, 81-90.

Peng R.D. (2002). *Multi-Dimensional Point Process Models in R*. Retrieved April 8, 2014, from http://www.biostat.jhsph.edu/~rpeng/papers/archive/ptprocR102.pdf

R CORE TEAM. R: a language and environment for statistical computing. The R Foundation for Statistical Computing, 2013, http://www.r-project.org/.

Rasmussen, J.G. (2013). Bayesian Inference for Hawkes Processes. *Methodology and Computing in Applied Probability*, *vol. 15*, 623-642.

Contact

Maciej Kostrzewski

Cracow University of Economics

ul. Rakowicka 27, 31-510 Kraków, Poland

maciej.kostrzewski@uek.krakow.pl