Tasmanian School of Business and Economics
University of Tasmania

Discussion Paper Series N 2017-06

A semi-parametric point process model of the interactions between equity markets

A.E. Clements

Queensland University of Technology, Australia

A.S. Hurn

Queensland University of Technology, Australia

K.A. Lindsay

Queensland University of Technology, Australia

V.V. Volkov

University of Tasmania, Australia

ISBN 978-1-86295-899-9

A semi-parametric point process model of the interactions between equity markets

A.E. Clements, A.S. Hurn, K.A. Lindsay
School of Economics and Finance, Queensland University of Technology

V.V. Volkov
Tasmanian School of Business and Economics, University of Tasmania

Abstract

A novel point process framework to examine the links between transaction data across equity markets is proposed. Moving beyond a simple exponential kernel specification, it is shown that the kernel matrix can be estimated by solving a system of integral equations which is uniquely characterised by second order cumulants. The cumulant based estimator is shown to be asymptotically normally distributed and consistent and is shown to perform well in a small simulation study. Applying this method to data from U.S. and U.K. equity markets when both are open, reveals that two-way interaction between trades is significant. Moreover, this interaction is characterised by both complex short term dynamics and long memory, which cannot be captured by conventional exponential kernels.

Keywords

Point processes, high-frequency data, conditional intensity

JEL Classification Numbers

C32, C51, C52, G10

Corresponding author

Vladimir Volkov
email: vladimir.volkov@utas.edu.au

1 Introduction

Point processes have been successfully applied to a wide range of event type data, and are a natural choice within which to consider high frequency financial market events. Since the seminal papers of Engle and Russell (1998), Russell (1999) and Engle (2000) point process models, usually variants of the self-exciting model proposed by Hawkes (1971a, 1971b), have been developed and applied to financial data. An important feature of these models is that the contemporaneous probability (or intensity) of an event in one point process is informed by the history of events in all processes via a kernel function. The kernel function provides a weighting of the relative importance of past events in determining the likelihood of a contemporaneous event in the process of interest.

Existing applications of point process models are predominantly based on an exponential kernel function (see, for example, Hautsch, 2011). The motivation for this choice is the property that in exponential kernels, the memory of past events decays at the same constant rate irrespective of when these events occur. There is, however, evidence to suggest that this property may not be universally desirable. Aït-Sahalia, Laeven and Pelizzon (2014) point out that clusters of large moves often do not occur exactly simultaneously across countries and that the transmission of the initial shock may take some time to be fully realised. Deo, Hurvich, Soulier and Wang (2009) show that traditional kernels cannot reproduce the long memory in the data that finally propagates into the realized volatility. Bacry and Muzy (2014) report that power-law kernels provide better performance than exponential kernels in modelling transactions data. These observations suggest strongly that kernels based on exponential decay may not always be an appropriate choice to model the complex dynamics in financial point process models for financial

transactions data.

This paper proposes a new approach to estimating the kernel function of a multivariate point process. It is shown that a multivariate mutually-exciting model can be characterised by first- and second-order sample cumulants. Furthermore, the kernel function of a mutually-exciting point process can be estimated by solving a system of integral equations based on the second-order cumulants, without the need for strong parametric assumptions about its shape. This approach has a number of benefits, the most obvious being that the absence of a parametric form for the kernel function allows for more complex dynamics. Moreover, obtaining the kernel by inverting the system of sample cumulants is computationally more robust than maximising a likelihood function. The proposed cumulant-based estimator is shown to be consistent and asymptotically normally distributed with a variance that only depends on the correlation structure of the point process. The finite sample performance of the estimator is demonstrated under simulation. Furthermore, it is shown that the maximum likelihood estimator of the kernel of a multivariate Hawkes model is a special case of the cumulant-based estimator, and consequently the latter possesses the same asymptotic properties as the former.

The development of the model is motivated by the desire to examine the interactions between transactions across different equity markets. A number of studies have applied Hawkes models to establish links between different types of high frequency events (trades and quotes for instance) for a particular asset (Bowsher, 2007; Large, 2007) or between the equity markets at a daily level (Aït-Sahalia, Cacho-Diaz and Laeven, 2015). The latter provides clear evidence that the United States and European markets are strongly interrelated. This paper focuses on the time period 14:30 - 16:30 GMT when

the stock markets in Chicago and London are both open for trading. During these hours trading volume is particularly high, and therefore this period can be expected to provide a wealth of information not only on trading activity but also on the interaction between these major international trading centres (see, for example, Hamao, Masulis and Ng (1990) who investigate price discovery). The primary finding is that the impact of Chicago trades on London trades is stronger than the influence in the reverse direction. Furthermore, a comparison of the semi-parametric kernel with its restricted exponential counterpart reveals that the interaction between markets is characterised by both complex short term dynamics and long memory, features that cannot be captured by the exponential kernel. This is a significant result which suggests that the standard Hawkes model based on an exponential kernel may underestimate cross-effects between the markets.

2 Modelling Intensity

Let $N = (N_1, \cdots, N_n)$ be an multivariate point process with associated counting variables $N(t) = (N_1(t), \cdots, N_n(t))$ that count the number of events in N occurring in $[0, t)$. Let N satisfy the following assumptions.

A1. The process N is orderly, so that

$$\mathbb{P}[N_i(t+h) - N_i(t) > 1] = o(h) \quad \forall i.$$

A2. The process N is weakly stationary.

A3. The process N is strongly mixing so that the distribution of the differential increments $dN_i(t)$ and $dN_j(s)$ behave as independent random variables for all values of i and j and values of t and s separated by a suitably large interval of time.

Given an n-variate mutually exciting point process a model for the conditional intensity of N_i is specified as

$$\lambda_i(t) = \mu_i + \int_0^\infty \sum_{j=1}^{n} G_{ij}(s) dN_j(t-s), \;\; i = 1, ..., n, \tag{1}$$

in which μ_i is an exogenous intensity, dN_j is the differential of the counting variable for process N_j and $G_{ij}(t)$ is the kernel function to be estimated from data. In other words, the contemporaneous intensity of the i-th process is given by a constant and its own past history, $G_{ii}(s)$, and the history of all the other processes, $G_{ij}(s)$ where $i \neq j$. The diagonal terms in $G_{jj}(s)$ are the self-exciting elements with the terms $G_{ji}(s)$ capturing cross-excitation when $i \neq j$. The process defined by $\lambda(t)$ is stationary if the spectrum of the matrix $\int_0^\infty G(s)ds$ lies within the unit ball.

The standard treatment of the kernel function is to specify an exponential form given by

$$G_{ij}(t) = \alpha_{ij} \exp(-\beta_{ij} t), \tag{2}$$

or to use Laguerre polynomials (Ogata, Akaike and Katsura, 1982; Ogata, 1988)

$$G_{ij}(t) = \sum_{k=0}^{m} \alpha_{ij}^{(k)} L_k(t) \exp(-\beta_{ij} t), \tag{3}$$

in which α_{ij} and β_{ij} are parameters of the kernels and Laguerre polynomials

$$L_k(t) = \sum_{s=0}^{k} \binom{k}{s} \frac{(-1)^s}{k!} t^s, \tag{4}$$

defined on $t > 0$.

To estimate the parameters governing the intensities in equation (1), the usual approach is to use maximum likelihood estimation applied to each intensity. The log-likelihood function for the i-th process is given by (see

Karr 1991, p.175)

$$\log L_i = \sum_{t=1}^{T} \Big[\log(\lambda_i)\, dN_i + (1 - dN_i) \log(1 - \lambda_i) \Big], \tag{5}$$

where parameters of the model appear implicitly via the value(s) of λ_i. Computation of the maximum likelihood parameter estimates is performed via the maximisation for each day d separately. In this case the likelihood function is given by

$$\log L_i(\theta_i) = \sum_{d=1}^{D} \sum_{t=1+T_{d-1}}^{T_d} \Big[\log(\lambda_{dt})\, dN_{dt} + (1 - dN_{dt}) \log(1 - \lambda_{dt}) \Big], \tag{6}$$

where D is the number of days and T_d represents the number of events to have occurred up to and including day d and $T_0 = 0$. Note that equation (6) implies that each day is an independent realisation of the same random process and overnight non-trading periods have no impact on intensity. The latter assumption can be relaxed by adding a link function to the intensity equation to capture the influence of overnight news (Bowsher, 2007).

The major advantage of the maximum likelihood approach is that when the kernel is parameterised in a relatively parsimonious way, for example by adopting the exponential specification (2) or the Laguerre specification (3) with a small number of polynomials m in the expansion, the procedure is straightforward to implement. Moreover, the maximum likelihood estimator of the intensity model inherits the desirable properties of all maximum likelihood estimators. For example, Ogata (1978) establishes that the maximum likelihood estimator for a simple, stationary, univariate point process model is consistent and asymptotically normal under certain regularity conditions.

While the parameters of the exponential and Laguerre specifications are easy to estimate, they may, however, not be an appropriate fit for the data. More appealing is the idea of an unrestricted (semi-parametric) memory

kernel. The difficulty is that the dimensionality of the parameter space rapidly overpowers a standard optimisation algorithm.

3 Estimation based on Cumulants

This section develops a method for estimating a semi-parametric kernel model based on the calculation of cumulants.

3.1 First- and Second-order Cumulants

The unconditional intensity of the process N_i is defined as

$$P_i = \frac{\mathbb{E}[dN_i(t)]}{dt}. \tag{7}$$

Assumptions A1-A3 allow $dN_i(t)$ to be interpreted in terms of expected values or probabilities (Brillinger, 1975). In other words, $P_i\, h$ is the unconditional probability of an event in $(t, t+h]$.

The joint intensity of events in N_i and N_j separated by an interval of duration u, also known as the second-order product density, $P_{ij}(u)$, is

$$P_{ij}(u) = \frac{\mathbb{E}[dN_i(t+u)dN_j(t)]}{dt\, du}. \tag{8}$$

In this case $P_{ij}(u)\, hk$ may be interpreted as the probability of an event in N_i during $(t, t+h]$ and simultaneously an event in N_j during $(t+u, t+u+k]$ for small values of h and k.

If events i and j satisfy the mixing condition A3, then increments become independent as $u \to \infty$ which means that

$$\lim_{|u|\to\infty} P_{ij}(u) = P_i P_j. \tag{9}$$

This suggests that a measure of the strength with which events in N_i and N_j separated by an interval u differ from independence is given by

$$Q_{ij}(u) = P_{ij}(u) - P_i P_j \,, \qquad u \neq 0 \,, \tag{10}$$

where $Q_{ij}(u)$ is known as the second-order cumulant. By virtue of assumption A3, the function $Q_{ij}(u) \to 0$ as $|u| \to \infty$. Moreover, it follows directly from the definition of the second-order product density that $Q_{ij}(-u) = Q_{j\,i}(u)$, so that the cumulant for negative lags is the transpose of the cumulant for positive lags.

The definition of the second-order cumulant can be extended to deal with $u = 0$. Simultaneity implies a discontinuity in Q_{ij}. By virtue of the fact that every process is simultaneous with itself at $u = 0$, then

$$Q_{ij}(u) = \left[\begin{array}{ll} P_{ij}(u) - P_i P_j & u \neq 0 \\ P_i \delta(u) & i = j \,, \quad u = 0 \,, \end{array} \right.$$

where $\delta()$ denotes the Dirac delta function. This methodology can be used to test the independence of a series of events. When two point processes N_i and N_j are independent, then $Q_{ij}(u) \equiv 0$. In particular, $Q_{ii}(u) = 0$ for a Poisson process when $u \neq 0$.

Suppose $t_r^{(i)}$ denote the times of events in the process N_i for a sample of duration T, then define

$$\Upsilon_{ij}^T(u) = \sum_{r,s} I\Big(\big| t_r^{(i)} - t_s^{(j)} - u \big| \leq \frac{\eta}{2} \Big),$$

where $I(\cdot)$ is the indictor function and the summation is taken over all relevant values of r and s. Cox (1965) shows that

$$\mathbb{E}[\Upsilon_{ij}^T(u)] \approx \eta T P_{ij}(u) \tag{11}$$

in which η denotes a window length. Using equation (11) the second-order product density is estimated by

$$\widehat{P}_{ij}(u) = \Upsilon_{ij}^{T}(u)/\eta T, \tag{12}$$

which together with equation (10) estimates the value of the cumulant at

$$\widehat{Q}_{ij}(u) = \Upsilon_{ij}^{T}(u)/\eta T - \widehat{P}_i\widehat{P}_j. \tag{13}$$

Theorem 1: Distribution of $\widehat{P}_{ij}^{1/2}(u)$
Given A1-A3 and fixed values of η, then

$$\sqrt{T}(\widehat{P}_{ij}^{1/2}(u) - P_{ij}^{1/2}(u)) \xrightarrow{d} \mathbb{N}\Big(0, (4T\eta)^{-1}\Big).$$

Moreover, $\widehat{P}_{ij}(u)$ and $\widehat{Q}_{ij}(u)$ are strongly consistent. ■

Proof: Brillinger (1976, 2012).

The importance of this result stems from the fact that the asymptotic variance of $\widehat{P}_{ij}^{1/2}(u)$ is independent of N. This property can be used to assign confidence intervals for use in ascertaining after what interval of time two processes N_i and N_j behave approximately independently. Thus plots of the square roots of product densities are useful for visualising periods of interdependence of processes, and therefore providing information as to how to set the interval of time over which the kernel function G_{ij} may be regarded as being nonzero.

3.2 Estimation

Estimating the kernel function using equation (1) starts by recognising that all contributions from lags $s > K$ are negligible. This results derives from the mixing condition, A3, satisfied by the component processes of N which

ensures that $Q(s) \to 0$ as $|s| \to \infty$, and also from the fact that $G(s) \to 0$ as $s \to \infty$. Therefore it is convenient to rearrange equation (1) into the form

$$Q(w) - G(w)\, D - \int_0^K G(s)\, Q(w-s)\, ds = \int_K^\infty G(s)\, Q(w-s)\, ds, \qquad (14)$$

in which $D = \mathrm{diag}(P_1, \cdots, P_n)$ is the $(n \times n)$ diagonal matrix of unconditional intensities. It is understood in subsequent analysis that the integral on the left hand side of equation (14) will be approximated by a quadrature, for example

$$\int_0^T f(x)\, dx = \sum_{j=0}^m \gamma_j f(s_j)\,, \qquad (15)$$

where $0 = s_0 < s_1 < \cdots < s_m = T$ are the nodes at which $f(x)$ is to be evaluated and $\gamma_0, \cdots, \gamma_m$ are the weights associated with these nodes. The right hand side of this equation will then assume the role of an error term with Frobenius norm satisfying

$$\begin{aligned}
\Big\| \int_T^\infty G(s)\, Q(w-s)\, ds \Big\|_F^2 &= \sum_{p,q=1}^n \Big(\int_T^\infty \sum_{i=1}^n G_{pi}(s)\, Q_{qi}(w-s)\, ds \Big)^2 \\
&\le \sum_{p,q=1}^n \Big(\int_T^\infty \Big[\sum_{i=1}^n G_{pi}^2(s) \Big]^{1/2} \Big[\sum_{i=1}^n Q_{qi}^2(w-s) \Big]^{1/2} ds \Big)^2 \\
&\le \sum_{p,q=1}^n \int_T^\infty \Big[\sum_{i=1}^n G_{pi}^2(s) \Big]\, ds \int_T^\infty \Big[\sum_{i=1}^n Q_{qi}^2(w-s) \Big]\, ds \\
&= \int_T^\infty \Big\| G(s) \Big\|_F^2 ds \int_T^\infty \Big\| Q(w-s) \Big\|_F^2 ds \\
&\le \int_T^\infty \Big\| G(s) \Big\|_F^2 ds \int_0^\infty \Big\| Q(s) \Big\|_F^2 ds\,.
\end{aligned}$$

Consequently, the error term can be made arbitrarily small provided the Frobenius norm of $Q(s)$ is square-integrable over $\mathbb{R}^+$ and that the convergence of $G(s)$ to zero as $s \to \infty$ is sufficiently strong so as to ensure that the Frobenius norm of G is square integrable over (T, ∞). Assuming these

conditions are satisfied, the kernel function $G(s)$ is ideally chosen to satisfy

$$h(w) = \left\| Q(w) - G(w)\, D - \int_0^T G(s)\, Q(w-s)\, ds \right\| = 0$$

for all $w \in (0, T)$. However, when the integral is approximated by the quadrature (15), the ideal objective is unachievable for all values of w, and instead the best outcome occurs when $G_0, \cdots, G_m$ are chosen to make $h(s_k) = 0$ for $k = 0, \cdots, m$, that is, $G_0, \cdots, G_m$ satisfy

$$Q(s_k) - G(s_k)\, D - \sum_{j=0}^{m} \gamma_j G(s_j)\, Q(s_k - s_j) = 0\,, \qquad k = 0, \cdots, m. \tag{16}$$

Estimation proceeds in the following way. First, use equation (13) to estimate the cumulant matrix $Q(t)$, then solve the system (14) to estimate the kernel $G(t)$. Finally, calculate the vector of exogenous intensities $\mu(t)$ using

$$\mu(t) = (\mathbb{I} - ||G(t)||)\Lambda(t), \tag{17}$$

in which $\mathbb{I}$ is the identity matrix, $||\cdot||$ denotes the L^1 norm of a matrix, and $\Lambda = \mathbb{E}[\lambda(t)]$. Deo, Hurvich and Lu (2006) propose an alternative way to estimate the deterministic component $\mu(t)$ using regression in the frequency domain. An alternative procedure for estimating the kernel function $G(t)$, proposed by Bacry, Dayri and Muzy (2012), requires the memory kernel to be symmetric, while in this paper this assumption in not required.

3.3 Asymptotic properties of the cumulant based estimator

Given an estimate of kernel $G(t)$, large sample properties of the cumulant based estimator can be investigated. First, equation (16) can be rewritten as

$$\begin{aligned} Q(t) &= G \circ (\delta(t)\mathbb{I} + Q(t)), \;\; t > 0, \\ &= G \circ \mathbb{Q}(t) \end{aligned} \tag{18}$$

in which $\mathbb{I}$ is the unit matrix and $\circ$ is the Fourier convolution. Equation (18) is analogous to the Yule-Walker system that relates kernels $G(t)$ with the cumulant matrices $Q(t)$ as

$$\Gamma_{u-1}[G(1),...,G(u)]' = [\widehat{Q}(1),...,\widehat{Q}(u)]', \tag{19}$$

$$\Gamma_{u-1} = \begin{pmatrix} \widehat{\mathbb{Q}}(0) & \dots & \dots & \widehat{\mathbb{Q}}(u-1) \\ \vdots & & & \vdots \\ \vdots & & \ddots & \vdots \\ \widehat{\mathbb{Q}'}(u-1) & \dots & \dots & \widehat{\mathbb{Q}}(0) \end{pmatrix}.$$

Under assumption A2, Γ_{u-1} is a non-singular matrix so that system (19) has a unique solution. In this case matrices $G(j)$ can be computed from Λ and the cumulant density matrices $Q(j)$.

Theorem 2: Consistency

Let N be a point process defined in equation (1) satisfying A1-A3. The cumulant based estimator of G, namely $\widehat{G}$, saisfies $\widehat{G}_{ij} \to G_{ij}$ in probability as $T \to \infty$. ■

Proof: Appendix

The asymptotic distribution of the cumulant based estimator of a mutually exciting process is given in the following theorem.

Theorem 3: Asymptotic normality

Given conditions A1-A3, a cumulant based estimator of a point process N is asymptotically distributed as

$$\sqrt{T}(\widehat{\mathbf{G}} - \mathbf{G}) \xrightarrow{d} \mathbb{N}\Big(\mathbf{0}_{s^2}, \Omega'\Sigma\Omega\Big),$$

in which $\mathbf{0}_s$ is a vector of s zeros, $\Sigma(i,j) = \mathrm{Cov}(\widehat{V}(i), \widehat{V}(j))$, $\widehat{V} = [\mathrm{vec}(\widehat{\Gamma}_{u-1}), \mathrm{vec}(\widehat{\mathbb{Q}}_u)]'$, vec is the column stacking operator of a matrix, $\widehat{\mathbf{G}} = \mathrm{vec}(\widehat{G})$, $\Omega' = -[\Gamma(1)\mathbf{I}_s, ..., \Gamma(s)\mathbf{I}_s]\left((\Gamma_s^{-1})' \otimes \Gamma_s^{-1}\right)[\mathbf{I}_{s^2} \;\; \mathbf{0}_{s^2 \times s}] + [\mathbf{0}_{s \times s^2} \;\; \Gamma_s^{-1}]$, $\otimes$ is the

Kronecker product, and $s = nu$. ■

Proof: Appendix

An important implication of Theorem 3 is that the asymptotic properties of the cumulant estimator $\widehat{\mathbf{G}}$ are defined by the first and second order characteristics of process N and a parametric specification of the kernel is not required. This result can be exploited to show asymptotic normality of maximum likelihood estimators as a special case. Let realisations of N be a sample from the k-dimensional *exponential family* with density

$$p_\theta(x) = c(\theta)a(x)\ \exp(\theta' t(x)), \tag{20}$$

where a and $t = (t_1, ..., t_k)$ are known functions on the sample space and θ is a parameter. The log-likelihood function $\ell_\theta(x) = \log\, p_\theta(x)$ can be differentiated (see e.g. Van der Vaart, 2000) with respect to θ. Hence, the likelihood equations $\sum \dot{\ell}_\theta(X_i) = 0$ can be represented as the system of k equations

$$\frac{1}{n}\sum_{i=1}^{s} t(X_i) = \mathbb{E}_\theta\, t(X).$$

Thus, the maximum likelihood estimators of kernels G are cumulant estimators, which is formalised in the following Lemma.

Lemma 1: Asymptotic properties of exponential families
Let N be a point process defined in equation (1), satisfying A1-A3, and $\widehat{\mathbf{G}}$ be the cumulant based estimator of $\mathbf{G}$. Asymptotic properties of the log-likelihood estimators of $\widehat{\mathbf{G}}$ depend on the function $e(\theta) = \mathbb{E}_\theta\, t(X)$. By differentiating $\mathbb{E}_\theta\, t(X)$ under the expectation sign, its derivative matrices are given by

$$\dot{e}_\theta = \mathrm{Cov}_\theta\, t(X),$$

in which the covariance matrices can be estimated from cumulants Q_{ij}. ■

Proof: Appendix

Note that the log-likelihood function, defined in equation (5) and parameterised by exponential kernels, is represented by a Poisson distribution which is part of the exponential family (20) and therefore asymptotic normality of estimates of a multivariate Hawkes model follows directly from Lemma 1. This extends the univariate results of Ogata (1978) to a multivariate case, namely that the maximum likelihood estimator for a simple, stationary, multi-variate point process model is consistent and asymptotically normal. Moreover, Theorem 3 and Lemma 1 confirm an unproved conjecture of Bowsher (2007) that the estimator of the Hawkes model is asymptotically normal.

4 Simulation Experiments

A simulation exercise was conducted in which the point process models with intensities

$$\lambda(t) = \mu + \alpha \int_0^\infty e^{-\beta s}\, dN(t-s), \tag{21}$$

$$\lambda(t) = \mu + \int_0^\infty \sum_{k=1}^{m} \alpha_k t^{k-1} \exp(-\beta t)\, dN(t-s) \tag{22}$$

were simulated and estimated. Each simulation exercise involved 200 independent replications of models (21) and (22) with parameter values $\mu = 0.001$, $\alpha = 0.003$ and $\beta = 0.02$ and $m = 5$, $\alpha_1 = 0.2$, $\alpha_2 = -0.0503$, $\alpha_3 = -0.0452$, $\alpha_4 = -0.0166$, $\alpha_5 = 0.0022$. With these parameter values, the simulated processes in (21) have mean rate 0.0012, or on average one event every 800 realisations, while the simulated processes (21) with the Laguerre kernel have mean rate 0.0010, or on average one event every 1000 realisations. Trials of length 250,000, 500,000, 1,000,000, 2,000,000 and 4,000,000 uniformly spaced realisations were run.

Table 1 reports the mean squared error (MSE) for the model mean in equations (21) and (22). The same sequences of trials are used to estimate the simulated kernels assuming a memory of 100 lags and applying the cumulant based procedure proposed earlier. As expected, the MSE for both models systematically decreases when the number of trials grow. The rate of convergence is consistent with the results reported in the previous section. Overall, the exponential kernel is estimated more accurately.

Table 1: The mean squared error of kernel estimates from the model with exponential and Laguerre kernels defined in equations (21) and (22). Values are multiplied by 10,000 and reported for 250,000, 500,000, 1,000,000, 2,000,000 and 4,000,000 independent trials.

	Exponential kernel	Laguerre kernel
$250,000$	0.0276	0.0533
$500,000$	0.0159	0.0233
$1,000,000$	0.0055	0.0152
$2,000,000$	0.0029	0.0071
$4,000,000$	0.0015	0.0032

While a simulated process with 4,000,000 trials has similar characteristics to models that will be discussed in the empirical section, it is interesting to visualise the respective kernels for this case. These kernels are plotted in Figure 1 at lag 100 with 5% confidence intervals. Only 2 estimated values for the Laguerre kernel and 4 values for the exponential kernel lie outside the confidence bounds, which confirms the accuracy of the proposed method.

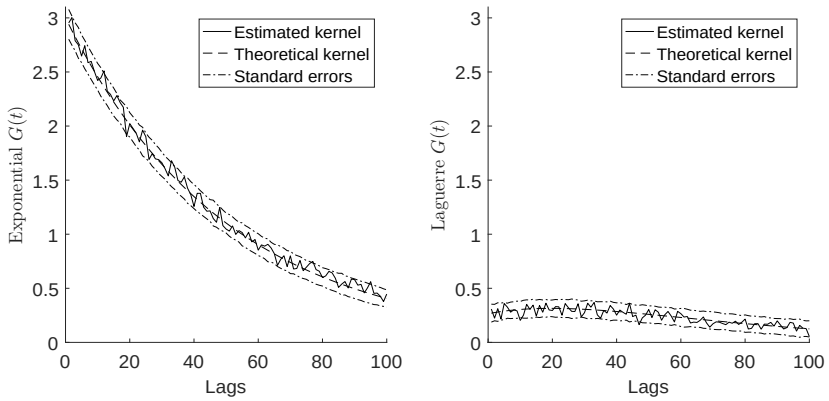


Figure 1: Kernel estimates for the simulated model defined in equations (21) and (22) with 4,000,000 independent trials were scaled by 1000. Confidence intervals are represented by $\pm 1.96 \times$ standard errors.

5 Data and preliminary analysis

Trading data for both the S&P 500 E-mini Index futures (traded on the Chicago Mercantile Exchange) and FTSE 100 Index futures (traded on the London Stock Exchange) contracts were downloaded from the Thomson Reuters Tick History database. Trades are recorded at millisecond timestamps with an associated trade price and volume. The data covers the period 3 January 2012 to 30 June 2012 for the times 14:30 to 16:25 (GMT), Monday to Friday, a period during normal trading hours when both markets are open and trading is active.

Consider Figure 2 which shows the average number of trades per minute for the period 14:30 to 16:25 (GMT). It is clear from Figure 2 that trading activity is higher around the opening of trading in Chicago (14.30 GMT), and around 15.00 GMT when most U.S. macroeconomic announcements occur, a finding consistent with Becker, Finnerty and Friedman (1995). However, the techniques proposed here to analyse the interaction between the markets

requires that the intensity of trading activity be relatively constant. Therefore only data from the period 15.10 to 16.10 GMT will be considered, as trading intensity in both markets is relatively constant during this period. Restricting the sample to this time period means that contamination from markets opening or macroeconomic announcements is avoided.

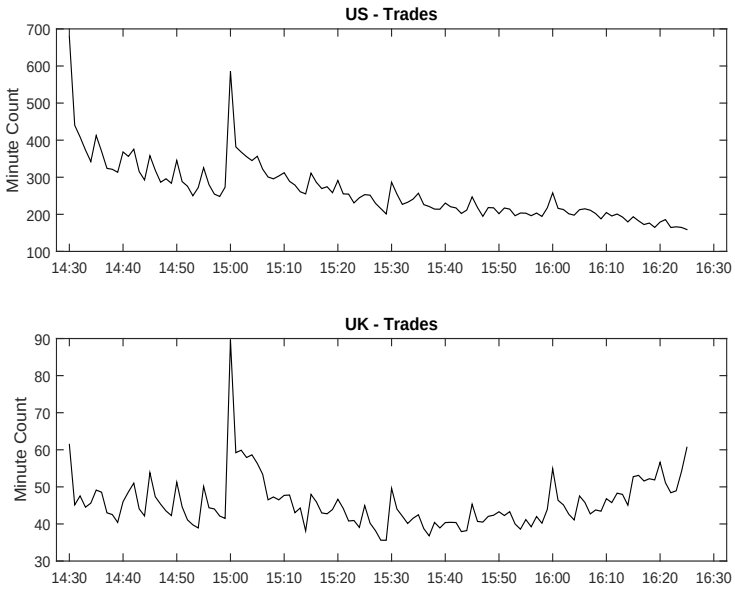


Figure 2: Number of trades per minute related to respective stock indices. Time in GMT.

The occurrence times of trades are mapped onto $(0,\infty)$ for each trading day in the same manner as Bowsher (2007). Two filters are applied to ensure the validity of the trade data. The ranges of the S&P 500 and FTSE 100 were taken to be $[1000, 2000]$ and $[5000, 8000]$ respectively in this investigation. A small number of trades with prices lying outside these range were removed. Additionally, a small number of trades with prices not at the minimum tick

size ($0.25 for the S&P 500 and £0.50 for the FTSE 100), or zero volume, or between the best bid-ask prices or with negative spreads were also removed from the final data set. Finally, trades with the same time stamp and price are treated as one event with an accumulated number of transactions.

		Trade durations
S&P 500	Number of durations	1533037
	Mean duration	258
	Standard deviation	469
	Minimum	1
	Maximum	13716
FTSE 100	Number of durations	279530
	Mean duration	1415
	Standard deviation	2283
	Minimum	1
	Maximum	38920

Table 2: Descriptive statistics for the durations between trade events for the S&P 500 and FTSE 100. The period is 3 January 2012 to 30 June 2012. The durations are measured in milliseconds.

Summary statistics of the durations between trades for the final data set are presented in Table 2. This shows that the intensity of trading in Chicago is greater and the average duration between trades in the S&P 500 contracts is more than 5 times shorter than that for the FTSE 100. In addition, both the volatility of the durations and the maximum duration is greater in the FTSE contracts. The product density estimator of Brillinger (1976) is used to examine the autocorrelation, or persistence in the trade point processes. This estimator is given by

$$\sqrt{\hat{P}_{ij}(u)} \sim \mathbb{N}\Big(\sqrt{P_{ij}(u)}, (4T)^{-1}\Big) \tag{23}$$

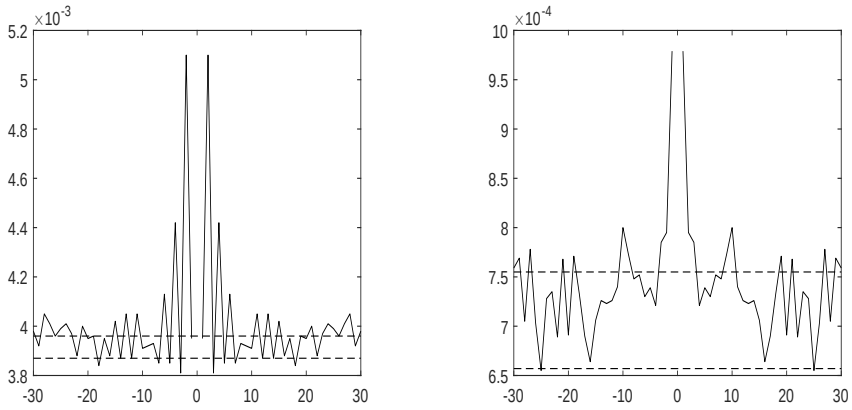
in which $P_{ij}(u)$ is the second order product density between events of type i and type j at a lag u and T is the sample size. The parameter estimates have a constant variance, which is used for setting confidence intervals to test the hypothesis of independent counting processes N_i and N_j. In this case, the 95% confidence limits for $\sqrt{\hat{P}_{ij}(u)}$ are

$$\sqrt{\hat{P}_i \hat{P}_j} \pm 1.96(4T)^{-1/2}, \tag{24}$$

in which $\hat{P}_i$ and $\hat{P}_j$ are the rates of the processes N_i and N_j respectively. If the estimated values lie inside the upper and lower confidence intervals, it can be considered evidence of the independence of the counting processes N_i and N_j. Plots of the square roots of product densities at lag u will identify the region of lags for which interactions between trades are significant.

The autocorrelations[1] in trades in both markets, represented by the estimates of the square root of product densities from equations (23) and (24), are shown in Figure 3. It is clear that S&P 500 trades are self-exciting exhibiting memory of around 10 minutes. The FTSE trades are less persistent, exhibiting memory out to around 5 minutes. Estimates of cross-correlation between the series of trades based on equation (23) are shown in Figure 4 and reveal a significant degree of dependence between the trades in each market. Positive (negative) lags in this case relate to S&P500 (FTSE) trades leading trades in the other market. Significant cross-correlations at positive and negative lags indicate that trades in both markets are strongly interrelated, an important result motivating a deeper investigation into the interaction between trades.

[1] In this section the term autocorrelation is used interchangeably for a lead-lag relationship.



(a) S&P 500 trades (b) FTSE 100 trades

Figure 3: Estimates of the square root of the product density from equation (23) for $T = 3600000$ are reported. The dashed horizontal lines are the estimated asymptotic values with the upper and lower 95% confidence limits under the hypothesis of independence. Lags are measured in minutes. Zero lags are excluded for illustration purposes.

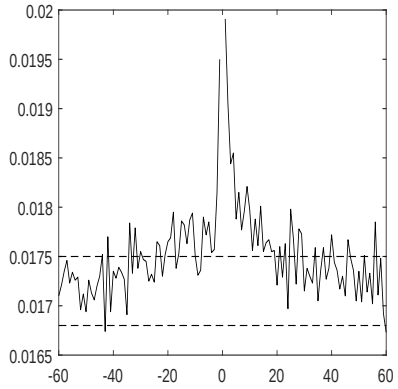


Figure 4: Estimates of the square root of the product density from equation (23) for $T = 3600000$ are reported. The dashed horizontal lines are the estimated asymptotic values with the upper and lower 95% confidence limits under the hypothesis of independence. A zero lag is excluded for illustration purposes. Lags are measured in seconds.

6 Empirical application

The preliminary analysis from the previous section shows that trades in both markets are strongly interrelated through complex short-term dynamics and long term memory. The main task is now to quantify this relationship without assuming an exponential memory kernel. This task is achieved by using the numerical procedure from Section 3 to estimate the multivariate conditional intensity model

$$\lambda(t) = \mu + \int_0^\infty G(s)\, d\mathrm{N}(t-s) \tag{25}$$

with

$$G(t) = \begin{bmatrix} G^s_{us} & G^c_{uk} \\ G^c_{us} & G^s_{uk} \end{bmatrix}, \tag{26}$$

in which $G^s_{us}(t)$ and $G^s_{uk}(t)$ capture the self exciting effects of trades in each market, while $G^c_{us}(t)$ and $G^c_{uk}(t)$ capture the cross excitation from the S&P500 and from the FTSE contracts respectively. A distinguishing feature of this model is its ability to capture bi-directional interactions between trades from each market.

Given a pair of records of occurrence times of S&P 500 and FTSE, $\{t_i \in I_{us}\}$ and $\{t_j \in I_{uk}\}$ respectively, and the exponential kernel specification in equation (2), the likelihood function in equation (5) can be rewritten as

$$\begin{aligned}
\log L_T(\theta) = & \sum_{t_i \in I_{us}} \log(\mu_{us} + \alpha^s_{us} R^s_{us}(i) + \alpha^c_{uk} R^c_{uk}(i)) \\
& - \mu_{us} T - \frac{\alpha^s_{us}}{\beta^s_{us}} \sum_{t_i \in I_{us}} (1 - \exp(-\beta^s_{us}(T - t_i)) - \frac{\alpha^c_{uk}}{\beta^c_{uk}} \sum_{t_j \in I_{uk}} (1 - \exp(-\beta^c_{uk}(T - t_j)) \\
& - \mu_{uk} T - \frac{\alpha^c_{us}}{\beta^c_{us}} \sum_{t_i \in I_{us}} (1 - \exp(-\beta^c_{us}(T - t_i)) - \frac{\alpha^s_{uk}}{\beta^s_{uk}} \sum_{t_j \in I_{uk}} (1 - \exp(-\beta^s_{uk}(T - t_j)) \\
& - \sum_{t_j \in I_{uk}} \log(\mu_{uk} + \alpha^c_{us} R^c_{us}(j) + \alpha^s_{uk} R^s_{uk}(j)),
\end{aligned} \tag{27}$$

where

$$R^s_{us}(i) = \exp(-(t_i - t_{i-1})\beta^s_{us})(1 + R^s_{us}(i-1)),$$

$$R^c_{us}(j) = \exp(-(t_j - t_{j-1})\beta^c_{us})(R^c_{us}(j-1)) + \sum_{i':t_{j-1}\le t_{i'}<t_j} \exp(-(t_j - t_{i'})\beta^c_{us}),$$

$$R^c_{uk}(i) = \exp(-(t_i - t_{i-1})\beta^c_{uk})(R^c_{uk}(i-1)) + \sum_{j':t_{i-1}\le t_{j'}<t_i} \exp(-(t_i - t_{j'})\beta^c_{uk}),$$

$$R^s_{uk}(j) = \exp(-(t_j - t_{j-1})\beta^s_{uk})(1 + R^s_{uk}(j-1)).$$

Table 3: Coefficient estimates of the Hawkes model (25). Parameter estimates (top) and t-statistics (bottom) are reported in each cell. Coefficients that are significant at the 5% level are marked (*).

	$\lambda_{\text{us}}(t)$		$\lambda_{\text{uk}}(t)$
$\hat{\alpha}^{\text{s}}_{\text{us}}$	0.0176^* (4.78)	$\hat{\alpha}^{\text{s}}_{\text{uk}}$	0.0059^* (4.61)
$\hat{\beta}^{\text{s}}_{\text{us}}$	0.0410^* (4.52)	$\hat{\beta}^{\text{s}}_{\text{uk}}$	0.0173^* (5.17)
$\hat{\alpha}^{\text{c}}_{\text{uk}}$	0.0073^* (3.28)	$\hat{\alpha}^{\text{c}}_{\text{us}}$	0.0079^* (5.72)
$\hat{\beta}^{\text{c}}_{\text{uk}}$	0.0216^* (3.62)	$\hat{\beta}^{\text{c}}_{\text{us}}$	0.0196^* (2.88)
$\hat{\mu}_{\text{us}}$	0.0022^* (4.58)	$\hat{\mu}_{\text{uk}}$	0.0005^* (3.77)

Coefficient estimates for the multivariate intensity model defined in equations (25) and (26), are presented in Table 3. All coefficients are significant, revealing complex interaction within, and between the two markets. The effect of self-excitation, captured by the coefficients α^s and β^s, is stronger in the S&P500 market. Moreover, the coefficient α^c_{us} and α^c_{uk}, capturing the cross effects between the Chicago and London equity markets, are very similar in magnitude.

In order to assess the impact of a trade event on the respective intensities, the estimated kernels from equation (26) are plotted in Figure 5. An interesting

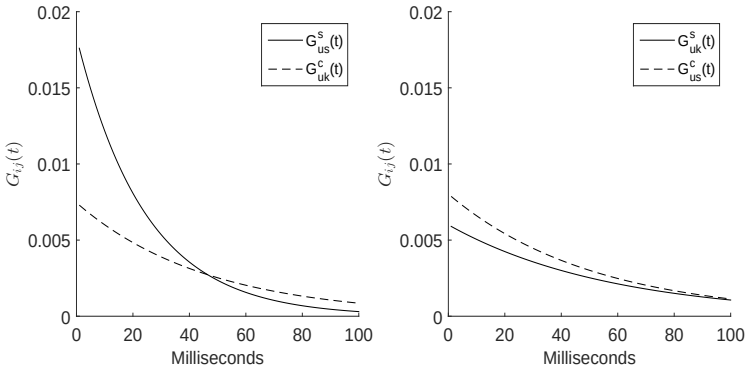


Figure 5: Exponential kernel estimates for the trade models defined in equations (25) and (26).

observation is that the effects from Chicago are stronger within the short intervals, while the impact from the London market is more significant at a longer horizon. This result confirms the finding of Huth and Abergel (2014) that the most liquid assets in terms of high trading turnover, which are the S&P500 contracts in this case, tend to lead the others at short horizons.

Results using the simple Hawkes model revels a significant interaction between the trades in both markets. In order to explore the question of whether or not the exponential kernel is too restrictive a specification, the semi-parametric kernel together with the cumulant-based estimation procedure will now be applied to the same data.

Consider the model defined by equation (25) without assuming the kernel function to be exponential. The kernel matrices $G(s)$ in equation (26) are now estimated using the algorithm described in Section 3. The semi-parametric kernel estimates are then compared with the estimates of the

Laguerre kernels[2]

$$G_{ij}(t) = \sum_{k=0}^{4} \alpha_{ij}^{(k)} L_k(t) \exp(-\beta_{ij} t). \tag{28}$$

Table 4: Coefficient estimates of the intensity model from equation (25) with Laguerre kernels from equation (28) are estimated using the maximum likelihood approach. Parameter estimates and t-statistics in parenthesis are reported in each cell. Calculations use $m = 4$ Laguerre polynomials.

	US $\leftrightarrow$ US	UK $\leftrightarrow$ UK	US $\rightarrow$ UK	UK $\rightarrow$ US
$\hat{\alpha}^{(0)}$	0.5142^*	0.0190^*	0.0071^*	0.0066^*
	(3.03)	(17.9)	(75.6)	(52.3)
$\hat{\alpha}^{(1)}$	-0.6703^*	-0.0037^*	0.0003^*	0.0001^*
	(4.52)	(9.40)	(10.7)	(4.18)
$\hat{\alpha}^{(2)}$	0.3461^*	0.0003^*	0.0001^*	0.0001^*
	(2.05)	(7.05)	(6.84)	(14.9)
$\hat{\alpha}^{(3)}$	-0.0756^*	-0.0001^*	-0.0001^*	-0.0001^*
	(2.04)	(5.32)	(11.6)	(15.3)
$\hat{\alpha}^{(4)}$	0.0059^*	0.0001^*	0.0001^*	0.0001^*
	(2.14)	(4.01)	(21.6)	(19.3)
$\hat{\beta}$	0.6144^*	0.0710^*	0.0564^*	0.0553^*
	(17.3)	(13.9)	(174)	(106)

The estimates of model (25) with the Laguerre kernels from equation (28) are presented in Table 4. The first and second columns of Table 4 contain estimates of the self-exciting kernels, while the third and fourth columns represent cross-exciting effects from Chicago and London respectively. All the coefficients are significant, which confirms that all five Laguerre polynomials capture dynamics of trading intensity that cannot be modelled by the simple exponential model.

[2]Ogata, Akaike and Katsura, (1982) give empirical arguments in favour of parametric Laguerre kernels rather than exponential kernels. For this reason, the performance of Laguerre kernels is compared with the semi-parametric estimates. The deterministic term μ is important for non-stationary models, where it can capture the diurnal patterns (see e.g. Engle and Russell, 1998; and Bowsher and Meeks, 2008). However, this term is not reported as the models discussed here deal with stationary data.

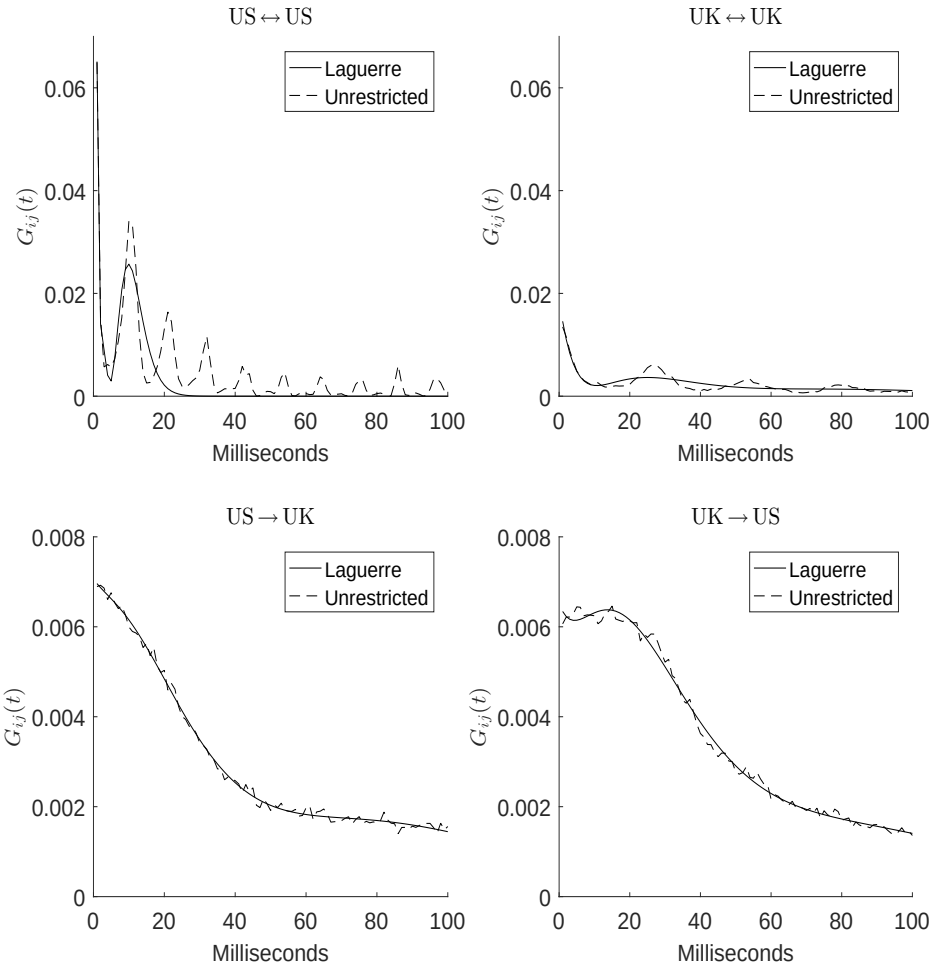


Figure 6: Kernel estimates for the trades model defined in equations (25) and (26). Effects from Chicago are presented in the first column and from London - second column.

The self-exciting effects reveal complex dynamics in both London and Chicago. In particular, the coefficients $\hat{\alpha}^{(0)}$ are positive and $\hat{\alpha}^{(1)}$ are negative, which means that in the short run the kernels decay and then rise in the longer

term. The cross-exciting effects are characterised by significant coefficients $\alpha^{(k)}$ supporting the conjecture discussed earlier that the kernels do not die out within 100 milliseconds.

To compare the Laguerre kernels from Table 4 with the corresponding semi-parametric kernels, both their estimates are presented in Figure 6. The semi-parametric self-exciting kernels show a periodic pattern: every 10 and 25 milliseconds there is a spike in the self-exciting responses. In the case of Chicago, both the Laguerre and unrestricted kernels decay then peak around 20 milliseconds. After this, the Laguerre decays to zero quickly whereas the unrestricted kernel exhibits periodic behaviour. The same pattern is evident in the case of self-excitation in London trading though the effect is not as pronounced. The cross-exciting kernels, however, exhibit quite different shapes. While the kernel capturing the effect of Chicago trading on the London market exhibits slow continual decay, the corresponding kernel function describing the impact of London trading on the Chicago market has a clear peak around 20 milliseconds, and thereafter exhibits continual decay almost identical to that of the former. In this case, both the Laguerre and unrestricted kernels exhibit very similar patterns. Overall, the more flexible kernels are capturing effects that are more complex than can be identified with simple exponential kernels. To compare the goodness of fit of the two kernels, residuals of the estimated model (25), defined as a rescaled point process after applying a random time change can be compared.[3] The sum of the squared residuals from the Laguerre kernel, 781.1, and the semi-parametric kernel, 329.4, show that the added flexibility of the unrestricted semi-parametric kernel is of benefit in explaining the dynamics of the trades in the two markets. Overall, these results show that complex

[3] A comparison of the residuals n unit-rate independent Poisson processes is described in Appendix E.

interactions between trades in Chicago and London are evident during the short period of overlap in trading. Given the methods proposed here, it is found that unrestricted semi-parametric kernels offer a better explanation of the dynamics in trading activity.

7 Conclusion

This paper proposes a novel approach for estimating multivariate point process models without restrictive assumptions regarding the shape of the memory kernel. The kernel of a point process can be estimated from data via the numerical solution of a system of integral equations. At its core, the strategy involves calculation of second-order cumulant functions from event data. Asymptotic properties of the proposed estimator are presented and its finite-sample performance investigated in a small simulation exercise. It is also shown that a standard maximum likelihood estimator of a multivariate Hawkes process has the same large sample properties as the cumulant based estimator.

As an empirical illustration of the efficacy of the new approach, an investigation of trade events for the time when both the Chicago and London markets are open is undertaken. The results show that the impact of Chicago trading on the London market is stronger than in the reverse direction. From a methodological perspective, one of the main conclusions to emerge is that the adoption of a memory function characterised by exponential decay is too restrictive. In this application at least, the semi-parametric estimators provides a superior fit to the short term dynamics and longer term memory of the trade interactions.

Appendices

A Asymptotic properties of product densities

The analysis begins with the assumption that $\widehat{P}_{NN}(u)$ is asymptotically Gaussian distributed with mean value $\mu = P_{NN}(u)$ and variance $\sigma^2 = (\beta T)^{-1} P_{NN}(u)$. Let $X = \widehat{P}_{NN}(u)$, then it is necessary to compute the asymptotic mean and variance of $\sqrt{X}$ given that $x \sim \mathbb{N}(\mu, \sigma^2)$.

Mean The mean value of $\sqrt{X}$ is by definition

$$\mathbb{E}[X^{1/2}] = \int_0^\infty \frac{\sqrt{x}}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}\, dx = \int_{-\mu}^\infty (z+\mu)^{1/2}\, \frac{1}{\sigma\sqrt{2\pi}}\, e^{-z^2/2\sigma^2}\, dz\,.$$

In the limit as $T \to \infty$ the integrand takes the limiting value $(z+\mu)^{1/2}\, \delta(z)$, where $\delta(z)$ is Dirac's Delta function. Thus $\mathbb{E}[\widehat{P}_{NN}(u)]$ tends asymptotically to $\sqrt{P_{NN}(u)}$.

Variance The variance of $\sqrt{X}$ is by definition

$$\begin{aligned}
\mathbb{V}[X^{1/2}] &= \int_0^\infty (\sqrt{x} - \sqrt{\mu})^2\, \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}\, dx \\
&= \int_{-\mu}^\infty (\sqrt{z+\mu} - \sqrt{\mu})^2\, \frac{1}{\sigma\sqrt{2\pi}}\, e^{-z^2/2\sigma^2}\, dz \\
&= \int_{-\mu}^\infty \frac{z^2}{(\sqrt{z+\mu} + \sqrt{\mu})^2}\, \frac{1}{\sigma\sqrt{2\pi}}\, e^{-z^2/2\sigma^2}\, dz\,.
\end{aligned}$$

The change of variable $z = \sigma\, y$ now gives

$$\beta\, T\, \mathbb{V}[X^{1/2}] = \frac{P_{NN}(u)}{\sqrt{2\pi}} \int_{-\mu/\sigma}^\infty \frac{y^2}{(\sqrt{\sigma\, y + \mu} + \sqrt{\mu})^2}\, e^{-y^2/2}\, dy\,.$$

Now let $T \to \infty$, or equivalently let $\sigma \to 0$ to obtain

$$\lim_{T\to\infty} \beta\, T\, \mathbb{V}[X^{1/2}] = \frac{P_{NN}(u)}{\sqrt{2\pi}} \int_{-\infty}^\infty \frac{y^2}{4\mu}\, e^{-y^2/2}\, dy = \frac{1}{4}$$

after noting that $\mu = P_{NN}(u)$, and completes the proof of the asymptotic properties of $P_{NN}^{1/2}(u)$.

B Theorem 2

A proof follows from the results of Theorem 1 about consistency of $P_{ij}(u)$ and $Q_{ij}(u)$.

C Theorem 3

As follows from Theorem 1, as $T \to \infty$ and $\beta = 1$,

$$\begin{bmatrix} \widehat{Q}(0) \\ \vdots \\ \widehat{Q}(u) \end{bmatrix} \sim \mathbb{N}\left(\begin{bmatrix} Q(0) \\ \vdots \\ Q(u) \end{bmatrix}, T^{-1}V_Q \right), \qquad (A1)$$

where V_Q is a matrix of respective product densities.

Consider the sample covariance matrix, $\widehat{\Gamma}_{u-1}$, and the matrix of second-order cumulants, $\widehat{\mathbb{Q}}_u$. Let

$$\widehat{V}_{\Gamma Q} = \begin{bmatrix} \mathrm{vec}(\widehat{\Gamma}_{u-1}) \\ \mathrm{vec}(\widehat{\mathbb{Q}}_u) \end{bmatrix} \qquad (A2)$$

be a vector with $n^2(u^2+u)$ elements. Then, using the result (A1), it is found that

$$\widehat{V}_{\Gamma Q} = \begin{bmatrix} \mathrm{vec}(\widehat{\Gamma}_{u-1}) \\ \mathrm{vec}(\widehat{\mathbb{Q}}_u) \end{bmatrix} \sim \mathbb{N}\left(\begin{bmatrix} \mathrm{vec}(\Gamma_{u-1}) \\ \mathrm{vec}(\mathbb{Q}_u) \end{bmatrix}, T^{-1}\Sigma_{\Gamma Q} \right), \quad (A3)$$

where $\Sigma_{\Gamma Q}$ is the $n^2(u^2+u) \times n^2(u^2+u)$ covariance matrix of $\widehat{V}_{\Gamma Q}$,

$$\Sigma_{\Gamma Q}(i,j) = \mathrm{Cov}(\widehat{V}_{\Gamma Q}(i), \widehat{V}_{\Gamma Q}(j)),$$

defined for $i, j = 1, ..., n^2(u^2 + u)$.

Now the Delta method (Van der Vaart, 2000) is used to demonstrate asymptotic normality of G. Let φ be the function from $\mathbb{R}^{n^2(u^2+u)}$ into $\mathbb{R}^{n^2u}$ defined by

$$\varphi(X) = (\text{unvec}_{n^2\times u^2+u}(X)C_1)^{-1}(\text{unvec}_{n^2\times u^2+u}(X)C_2),$$

where X is a vector with $n^2(u^2 + u)$ elements,

$$C_1 = \begin{bmatrix} \mathbf{I}_{nu} \\ \mathbf{0}'_{n\times nu} \end{bmatrix}, \qquad C_2 = \begin{bmatrix} \mathbf{0}_{nu\times n} \\ \mathbf{I}_n \end{bmatrix},$$

are a $(nu + n) \times nu$ matrix and a matrix with $n^2(u + 1)$ elements, respectively, $\mathbf{0}_{p\times q}$ is a matrix of zeros and $\text{unvec}_{p\times q}(X)$ is the inverse of the vec operator defined such as $\text{unvec}_{p\times q}(\text{vec}(A)) = A$ (Further, the subscript of this operator is omitted for simplicity).

Note that for $\widehat{V}_{\Gamma Q}$ defined in (A2)

$$\varphi(\widehat{V}_{\Gamma Q}) = \widehat{\Gamma}^{-1}_{u-1}\widehat{Q}_u = \widehat{G}.$$

Then, by the Delta method,

$$\widehat{\mathbf{G}} \sim \mathbb{N}(\mathbf{G}, T^{-1}D'\Sigma_{\Gamma Q}D),$$

where $\widehat{\mathbf{G}} = \text{vec}(\widehat{G})$ and

$$D = \left(\frac{\partial\varphi}{\partial X}\right)_{|X=\widehat{V}_{\Gamma Q}} = \left(\frac{\partial\varphi}{\partial X'}\right)'_{|X=\widehat{V}_{\Gamma Q}}$$

is the $n^2(u^2+u)\times n^2u$ derivative matrix of the function φ, defined as follows. Let $f : \mathbb{R}^p \to \mathbb{R}^q$ be a vector valued function with vector variable. The $p \times q$ matrix derivative of f is defined as

$$\frac{\partial f}{\partial X} = \left(\frac{\partial f}{\partial X'}\right)' = \begin{bmatrix} \frac{\partial f_1}{\partial X_1} & \cdots & \frac{\partial f_1}{\partial X_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_q}{\partial X_1} & \cdots & \frac{\partial f_q}{\partial X_p} \end{bmatrix}',$$

in order to meet the needs of the Jacobian matrix.

Furthermore, $\varphi(X)$ can be represented as the product of two matrices, $\varphi(X) = M_1^{-1}(X)M_2(X)$. Then, by using the product rule, it is found that

$$\begin{aligned}\left(\frac{\partial \varphi}{\partial X'}\right)' &= \left(\frac{\partial (M_1^{-1}M_2)}{\partial X'}\right)' \\ &= \left((M_2' \otimes \mathbf{I}_{nu})\frac{\partial \text{vec}(M_1^{-1})}{\partial X'} + (\mathbf{I}_1 \otimes M_1^{-1})\frac{\partial \text{vec}(M_2)}{\partial X'}\right)'.\end{aligned}$$

Now, using the matrix derivative of an inverse matrix,

$$\frac{\partial \text{vec}(M_1^{-1})}{\partial X'} = -\left((M_1^{-1})' \otimes M_1^{-1}\right)\frac{\partial \text{vec}(M_1)}{\partial X'},$$

replacing M_1 by $\text{unvec}(X)C_1$, and using the properties of vec and unvec operators (Neudecker, 1969) and the derivative matrix rule of f such that $f = AX$, where A is a constant matrix, it is found that

$$\begin{aligned}\frac{\partial \text{vec}(M_1^{-1})}{\partial X'} &= -\left(((\text{unvec}(X)C_1)^{-1})' \otimes (\text{unvec}(X)C_1)^{-1}\right)\frac{\partial \text{vec}(\text{unvec}(X)C_1)}{\partial X'} \\ &= -\left(((\text{unvec}(X)C_1)^{-1})' \otimes (\text{unvec}(X)C_1)^{-1}\right)\frac{\partial (C_1' \otimes \mathbf{I}_{nu})X}{\partial X'} \\ &= -\left(((\text{unvec}(X)C_1)^{-1})' \otimes (\text{unvec}(X)C_1)^{-1}\right)(C_1' \otimes \mathbf{I}_{nu}).\end{aligned}$$

Applying the same rule to M_2,

$$\frac{\partial \text{vec}(M_2)}{\partial X'} = \frac{\partial (C_2' \otimes \mathbf{I}_{nu})X}{\partial X'} = (C_2' \otimes \mathbf{I}_{nu}).$$

Then,

$$\begin{aligned}\left(\frac{\partial \varphi}{\partial X'}\right)' = \Big(&- ((\text{unvec}(X)C_2)' \otimes \mathbf{I}_{nu}) \\ &(((\text{unvec}(X)C_1)^{-1})' \otimes (\text{unvec}(X)C_1)^{-1})(C_1' \otimes \mathbf{I}_{nu}) \\ &+ (\mathbf{I}_{nu} \otimes (\text{unvec}(X)C_1)^{-1})(C_2' \otimes \mathbf{I}_{nu})\Big)'.\end{aligned}$$

Therefore,

$$\begin{aligned} D^{'} &= -(\gamma^{'} \otimes \mathbf{I}_{nu})((\Gamma_{nu}^{-1})^{'} \otimes \Gamma_{nu}^{-1})(C_1^{'} \otimes \mathbf{I}_{nu}) + (\mathbf{I}_1 \otimes \Gamma_{nu}^{-1})(C_2^{'} \otimes \mathbf{I}_{nu}) \\ &= -\left[\Gamma(1)\mathbf{I}_s, ..., \Gamma(s)\mathbf{I}_s\right] \left((\Gamma_s^{-1})^{'} \otimes \Gamma_s^{-1}\right) \left[\mathbf{I}_{s^2} \;\; \mathbf{0}_{s^2 \times s}\right] + \left[\mathbf{0}_{s \times s^2} \;\; \Gamma_s^{-1}\right] , \end{aligned}$$

where $\gamma = \text{vec}(\mathbf{Q})$, $s = nu$ and the theorem follows.

D Lemma 1

As follows from Lemma 4.5 of Van der Vaart (2000), the vector of partial derivatives (the score function) satisfies

$$\dot{\ell}_\theta(x) = \frac{\dot{c}}{c}(\theta) + t(x) = t(x) - \mathbb{E}_\theta t(X).$$

The second equality shows that score functions have zero means. It can be formally established by differentiating the identity $\int p_\theta ds = 1$ under the integral sign. Combining the Lemma 4.5 of Van der Vaart (2000) and the Leibniz rule gives

$$\frac{\partial}{\partial \theta_i} \int p_\theta ds = \int \frac{\partial c(\theta)}{\partial \theta_i} a(x) \exp(\theta^{'} t(x)) ds(x) + \int c(\theta) a(x) t_i(x) \exp(\theta^{'} t(x)) ds(x).$$

The left side is zero and the equation can be rewritten as $0 = \dot{c}/c(\theta) + \mathbb{E}_\theta t(X)$, which proves the lemma.

E A goodness of fit test

In order to perform a goodness of fit test of the estimated model (25) one can use the following corollary 14.6.V. of Daley and Vere-Jones (2008):

Corollary 1 *(Random time change[4]) Let N denote a nonterminating multivariate point process with components $N_i, i = 1, ..., n$ and $\mathfrak{F}$-conditional*

[4] See Daley and Vere-Jones (2008) for the proof.

intensity $\lambda_i(t)$. *Let* $a(t,i) = \int_0^t \lambda_i(s)ds$ *and denote by* $\tilde{N}$ *the rescaled point process defined to have a point at* $\{a(t,i), i\}$ *if and only if the* i*th component of* N *contains a point at* t. *Then* $\tilde{N}$ *is a stationary compound Poisson process with unit intensity.*

Equivalently, if the rescaling is performed so that $\tilde{N}$ *has a point at* $(a(t,i), i)$ *whenever the original process has a point at* (t,i), *then the resultant process consists of* n *independent, unit-rate Poisson processes.*

The corollary 1 allows to transform a point process N to n unit-rate independent Poisson processes. This result can be applied to the pooled process received by superposing all events for two markets. The residuals of this process can be defined as

$$e_j^p = \int_{t_j}^{t_{j+1}} \sum_{i=1}^{2} \lambda_i(s)ds, \qquad (j = 0, 1, ...),$$

where e_j^p is the Poisson process with intensity 2. Given the estimates of kernel $G_{ij}(\cdot)$ and exogenous intensity μ, the rescaled residuals[5] e_j^p can be plotted in the unite square with the confidence lines of the standard normal distribution $\pm Z_{1-\alpha/2}/\sqrt{T}$, in which T is the length of the sample. If the empirical values e_j^p falls outside the confidence bands it is an evidence of unsatisfactory fit.

Given the estimates $G_{ij}(\cdot)$ of the model (25), a goodness of fit test of the model with Laguerre and semi-parametric kernels was conducted. The goodness of fit results indicate a poor performance of the model with the Laguerre kernel comparing to the semi-parametric model.

[5]Before plotting the residuals e_j^p should be multiplied by the intensity of the pooled process, which is equal to 2.

References

Aït-Sahalia, Y., Laeven, R.J.A., and Pelizzon, L. 2014. Mutual excitation in eurozone sovereign CDS. *Journal of Econometrics*, **183**, 151–167.

Ait-Sahalia, Y., Cacho-Diaz, J., and Laeven, R.J.A. 2015. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, **117**, 585–606.

Bacry, E., and Muzy, J.F. 2014. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, **14**, 1147–1166.

Bacry, E., Dayri, K., and Muzy, J.F. 2012. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems*, **85**, 1–12.

Becker, K.G., Finnerty, J.E., and Friedman, J. 1995. Economic news and equity market linkages between the US and UK. *Journal of Banking & Finance*, **19**, 1191–1210.

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

Bowsher, C.G., and Meeks, R. 2008. The dynamics of economic functions: modeling and forecasting the yield curve. *Journal of the American Statistical Association*, **103**, 1419–1437.

Brillinger, D.R. 1975. The identification of point process systems. *The Annals of Probability*, **3**, 909–924.

Brillinger, D.R. 1976. Estimation of the second-order intensities of a bivariate stationary point process. *Journal of the Royal Statistical Society. Series B (Methodological)*, **52**, 60–66.

Brillinger, D.R. 2012. *Statistical inference for stationary point processes.* Springer.

Cox, D.R. 1965. On the estimation of the intensity function of a stationary point process. *Journal of the Royal Statistical Society. Series B (Methodological)*, **27**, 332–337.

Daley, D.J., and Vere-Jones, D. 2008. *An introduction to the theory of point processes: General theory and structure.* Vol. 2. Springer Verlag.

Deo, R., Hurvich, C., and Lu, Y. 2006. Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics*, **131**, 29–58.

Deo, R., Hurvich, C.M., Soulier, P., and Wang, Y. 2009. Conditions for the propagation of memory parameter from durations to counts and realized volatility. *Econometric Theory*, **25**, 764–792.

Engle, R.F. 2000. The Econometrics of ultra-high-frequency data. *Econometrica*, **68**, 1–22.

Engle, R.F., and Russell, J.R. 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127–1162.

Hamao, Y., Masulis, R.W., and Ng, V. 1990. Correlations in price changes and volatility across international stock markets. *Review of Financial Studies*, **3**, 281–307.

Hautsch, N. 2011. *Econometrics of financial high-frequency data.* Springer Verlag.

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

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

Huth, N., and Abergel, F. 2014. High frequency lead lag relationships - empirical facts. *Journal of Empirical Finance*, **26**, 41–58.

Karr, A. 1991. *Point processes and their statistical inference.* Baltimore, Maryland: CRC press.

Large, J. 2007. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, **10**, 1–25.

Neudecker, H. 1969. Some theorems on matrix differentiation with special reference to Kronecker matrix products. *Journal of the American Statistical Association*, **64**, 953–963.

Ogata, Y. 1978. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, **30**, 243–261.

Ogata, Y. 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, **83**, 9–27.

Ogata, Y., Akaike, H., and Katsura, K. 1982. The application of linear intensity models to the investigation of causal relations between a point

process and another stochastic process. *Annals of the Institute of Statistical Mathematics*, **34**, 373–387.

Russell, J.R. 1999. Econometric modeling of multivariate irregularly-spaced high-frequency data. *Manuscript, GSB, University of Chicago.*

Van der Vaart, A.W. 2000. *Asymptotic statistics.* Cambridge University Press.