# Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models[*]

Clive G. Bowsher[†]

Nuffield College, University of Oxford, Oxford OX1 1NF, U.K.

5 November 2002

## Abstract

A continuous time econometric modelling framework for *multivariate* market event (or 'transactions') data is developed in which the model is specified via the vector stochastic intensity. This has the advantage that the conditioning $\sigma$-field is updated continuously in time as new information arrives. We introduce the class of generalised Hawkes models which allow the estimation of the dependence of the intensity on the events of previous trading days. Analytic likelihoods are available and we show how to construct diagnostic tests based on the transformation of non-Poisson processes into standard Poisson processes using random changes of time. A proof of the validity of the diagnostic testing procedures is given that imposes only a very weak condition on the point process model, thus establishing their widespread applicability. A continuous time bivariate point process model of the timing of trades and mid-quote changes is presented for a NYSE stock and the empirical findings are related to the theoretical and empirical market microstructure literature.

*Keywords:* Point and counting processes, intensity, multivariate, diagnostics, goodness of fit, specification tests, change of time, transactions data, NYSE, NASDAQ, market microstructure

## 1 Introduction

The availability of so-called 'ultra-high-frequency' or 'transactions' data in empirical finance has resulted in considerable interest by econometricians in the development of models to analyse the intraday behaviour of financial markets. This data records the timing and characteristics of all occurrences of certain types of market event during the trading day.[1] For example, the New York Stock Exchange (NYSE) Trade and Quote (TAQ) database reports the timing, price, and volume of all reported trades, together with the timing and details of all changes to the market quotes. Such datasets make possible an unprecedented view of the detailed workings

---

[1]Henceforth, we refer to such data as 'market event data' because the term 'transactions' is often taken to be synonymous with trades.

of financial markets and allow researchers to evaluate the hypotheses of the theoretical market microstructure literature.

The particular contribution of this paper is the development of a continuous time econometric modelling framework for market event data that can readily be applied to the *multivariate* case. By multivariate, we mean that more than one *type* of event is considered – for example, in the bivariate case, trades and changes to the market quotes, or trades in two related assets. A relevant issue to be considered at this point is whether the obvious difficulties of modelling multivariate market event data could not be mitigated somewhat by adopting a simpler approach based on time series methods, rather than attempting to set the models in continuous time. Indeed, as Hasbrouck (1999) points out, these difficulties have strongly motivated simpler approaches such as specifying a time series model in 'event time' where $t$ indexes trades (see, for example, Hasbrouck (1991)) or aggregating data over fixed intervals of real time (as in Hasbrouck (1999)).

We believe the development of statistical models for market event data set in continuous, real time to be an important challenge in financial econometrics for the following reasons. First, models set in event time may well ignore aspects of the evolution of the market that are economically important. Indeed, a growing number of papers point to the economic importance of 'wall-clock' time. Easley and O'Hara (1992) establish in a sequential trade framework that when there is uncertainty over whether an information event has occurred, the time intervals between trades are informative and so affect the quoted prices. Dufour and Engle (2000) provide empirical evidence that as the intensity of trading in real time increases, the price impact of trades, the speed of price adjustment to trade related information and the positive autocorrelation of signed trades increase. In addition, Hasbrouck (1999) examines empirically the issue of time deformation in equity markets and concludes that the roles of trades, orders and quotes are different in 'fast' and 'slow' markets (where these terms are taken to refer to the evolution of the market in real time). Such papers highlight the limitations of models that ignore real time altogether. In addition, most practical applications of models of market event data such as volatility measurement and the design of optimal order submission strategies (see Harris (1998)) require that the models relate somehow to real time.

A standard time series analysis of aggregated data using fixed intervals of real time is also problematic. Since the data records the timing and characteristics of individual market events, aggregation involves an undesirable loss of information. As Engle and Russell (1998) warn, "the characteristics and timing relations of individual transactions will be lost, mitigating the advantages of moving to transactions data in the first place." The considerations set out above suggest that models for market event data set in continuous time are likely to provide important economic insights into the functioning of financial markets.

Econometrically speaking, market event data can be viewed as the realisation of a Marked Point Process (hereafter Marked PP or MPP): that is, as the realisation of a double sequence $\{T_i, Z_i\}_{i \in \{1,2,...\}}$ of random variables where $T_i$ is the random occurrence time of the $i$th event and $Z_i$ is a vector of additional variables (or 'marks') associated with that event. We develop an econometric modelling framework for multivariate financial PP data – that is, for data generated by a MPP where $Z_i \in \{1, 2, ..., M\}$ indicates the type of the $i$th event. The distinguishing feature of our approach is that the model is specified via the vector stochastic intensity rather than directly specifying a model for the durations (or time intervals) between events. This is particularly advantageous when a multivariate PP model or a model for a univariate PP conditional on some other continuous time process is needed. Consider, for example, a model for the timing of mid-quote changes for a NYSE stock where we condition on the timing of trades. Specifying the model via the mid-quote intensity rather than via the conditional distribution of durations between mid-quote changes allows the intensity to change immediately in response to the occurrence of a trade during a duration. Because the filtration is updated continuously in time and includes the history generated by the trade arrival process, the additional information since the start of the duration can be taken into account.[2] Whilst considerable progress has been made in modelling the univariate case using time series models of durations (most notably the Autoregressive Conditional Duration models of Engle and Russell (1997,1998)), multivariate extensions of this work have been slow to emerge in the econometrics literature.[3] One such

---

[2]Hamilton and Jordà (2001) use an intensity based approach to model a discrete time point process and also note the advantage that this offers in terms of being able to incorporate information that occurs after the most recent event has taken place.

[3]Univariate models of market event data, and ACD models in particular, are surveyed by Bauwens and Giot (2001).

example is the model of Engle and Lunde (2003) for trades and quotes which is discussed in Section 6.1 towards the end of the paper.

We believe that approaching the problem of modelling multivariate market event data by directly specifying the stochastic intensity provides a powerful, flexible framework. We introduce here a family of models – the generalised Hawkes models – for which analytic likelihoods are available and diagnostic tests based on the integrated intensity can be constructed.[4] In contrast to previous work, the models are general enough to allow us to estimate the nature of the dependence of the intensity on the events of previous trading days rather imposing strong *a priori* assumptions concerning this dependence. This approach recognises a significant feature of financial markets: namely, that for the majority of markets, the market does not operate continuously so that the question of dependence between trading days is of considerable importance. Furthermore, the models have intuitively appealing economic interpretations since the stochastic intensity for events of a particular type can be interpreted as a conditional hazard – that is, as the conditional expectation in the limit of the number of events of that type that will occur per unit time in the future given the current (multivariate) information set. Our approach is closest to that of Russell (1999) who also specifies a multivariate PP model via the stochastic intensity. A comparison with his modelling framework is provided in Section 6.2.

The structure of the paper is as follows. Section 2 provides a short introduction to the modern intensity theory of PPs and makes readily accessible those aspects of the PP literature that are particularly relevant to the econometric analysis of market event data. With the advent of high frequency intraday financial data, PPs – although relatively unfamiliar – are becoming an important tool for financial econometricians. We have therefore devoted some space to the careful definition of central PP concepts and to the discussion of statistical inference for the (broad) class of PP models where the stochastic intensity is available in closed form. Section 3 builds on the suggestions of Russell (1999) for the construction of diagnostic tests in this context. We show that the tests can be thought of as being based on the transformation of non-Poisson processes into standard Poisson processes using a random change of time, and thus establish that

---

[4]We use the term diagnostic test in the sense of a specification test. The term goodness of fit test would be more usual in the statistical literature.

the PP model need only satisfy a very weak condition indeed in order for the testing procedures to be valid. Various diagnostic tests are proposed, some of which are new to the econometrics PP literature. Section 4 presents the basic 'building blocks' of the new models in the simpler univariate setting before the discussion is extended to the bivariate case in Section 5. Both Sections 4 and 5 contain empirical applications using data from the TAQ database and illustrate the empirical usefulness of generalised Hawkes processes. In particular, Section 5 presents a continuous time bivariate PP model of the timing of trades and mid-quote changes for a NYSE stock and relates the empirical findings to the theoretical and empirical market microstructure literature. Section 6 explores connections between our work and previous research, and discusses possible extensions – in particular, modelling the evolution of the multivariate PP conditional on a wider filtration than its internal history. Section 7 concludes. The Appendix describes the adjustments that are made to the TAQ data and presents the results of a sensitivity analysis that establishes the robustness of our empirical findings in Section 5 to the particular adjustment rule that is employed. The Appendix also discusses how to numerically compute maximum likelihood estimates for the models.

## 2  Point Processes and Stochastic Intensities

This section describes how a statistical model for (possibly multivariate) PP data can be completely specified via a parametric family of stochastic intensities and establishes that a legitimate likelihood function and likelihood ratios are available. In so doing, we provide definitions of an $M$-variate PP, its associated counting process $N(t)$, and its internal history $\mathcal{F}_t^N$. Theorem 2.1 expresses the likelihood in terms of the stochastic intensity. Finally, known asymptotic results concerning the maximum likelihood estimator (MLE) and likelihood ratio (LR) test for stationary PP models are stated and discussed. Textbook treatments of the modern, martingale-based intensity theory of PPs are given in Brémaud (1981) and Karr (1991), from which we draw extensively in this section.

## 2.1 Point and Counting Processes

**Definition 1 *Point Process.*** *Let $\{T_i\}_{i \in \{1,2,...\}}$ be a sequence of nonnegative random variables on some probability space $(\Omega, \mathcal{F}, P)$ such that $0 < T_i \leq T_{i+1}$. Then the sequence $\{T_i\}$ is called a point process on $[0, \infty)$. If in addition, $T_i < T_{i+1}$ $\forall i$ then the point process is said to be a simple point process.*

**Definition 2 *M-variate point process and counting process.*** *Let $\{T_i\}_{i \in \{1,2,...\}}$ be a simple point process on $[0, \infty)$, defined on $(\Omega, \mathcal{F}, P)$, and let $\{Z_i\}_{i \in \{1,2,...\}}$ be a sequence of $\{1, 2, ..., M\}$-valued random variables (also defined on $(\Omega, \mathcal{F}, P)$, with $1 \leq M < \infty$). Then the double sequence $\{T_i, Z_i\}_{i \in \{1,2..\}}$ is called an M-variate point process on $[0, \infty)$. Define for all $m, 1 \leq m \leq M$, and all $t \geq 0$*

$$N_m(t) = \sum_{i \geq 1} 1(T_i \leq t)1(Z_i = m). \tag{1}$$

*Then the M-vector process $N(t) = (N_1(t), ..., N_M(t))$ is the M-variate counting process associated with $\{T_i, Z_i\}$.[5]*

In our context, $T_i$ will be the occurrence time of the $i$th market event and $Z_i$ will indicate the event's type.[6] $N_m(t)$ gives the random number of events of type $m$ that have occurred up to and including time $t$.[7] Because $\{T_i\}$ in the above definition is simple, the possibility of the simultaneous occurrence of two events (of either the same or different types) is ruled out.

Some comments on notation follow. Let $X(t)$ be a measurable process whose sample paths are either locally bounded or nonnegative. Then the *stochastic Stieltjes integral* of $X$ with respect to $N_m$ exists, and for each $t$,

$$\int_{(0,t]} X(s)dN_m(s) = \sum_{i \geq 1} 1(T_i^{(m)} \leq t)X(T_i^{(m)}), \tag{2}$$

---

[5]We allow for the univariate case ($M = 1$) in the definition of an $M$-variate PP as this simplifies the statement of some of the theorems that follow.

[6]$Z_i$ is sometimes referred to as the mark of the $i$th event. Indeed, an $M$-variate PP is a special case of a Marked Point Process, $\{T_i, Z_i\}_{i \in \{1,2..\}}$, in which the mark space is the finite set $\{1, 2, ..., M\}$.

[7]In what follows we sometimes refer, in a slight abuse of terminology, to 'the $M$-variate PP $N(t)$.' This is harmless since there is a one-one mapping between the $M$-variate point and counting processes for all $M \geq 1$. When $M > 1$, we also sometimes refer to $\{T_i\}$ (or, equivalently, $\sum_{m=1}^{M} N_m(t)$) as the 'pooled process' since each of its realisations is formed by pooling events of all types.

where $T_i^{(m)}$ denotes the occurrence time of the $i$th $m$-type event.[8] For each $t$ and $\omega$, the integral in (2) is the usual Lebesgue-Stieltjes integral. The *internal history* of the $M$-variate PP $N(t)$ is denoted by $\{\mathcal{F}_t^N\}_{t\geq 0}$, where $\mathcal{F}_t^N = \sigma(N_A(s) : 0 \leq s \leq t, A \in \mathcal{E})$, $N_A(s) = \sum_{i\geq 1} 1(T_i \leq s)1(Z_i \in A)$, and $\mathcal{E}$ is the $\sigma$-field of all subsets of $\{1, 2, ..., M\}$.[9] $\mathcal{F}_t^N$ corresponds to complete observation of the $M$-variate PP up to (and including) time $t$. Also $\mathcal{F}_{T_i}^N := \sigma(T_1, Z_1, T_2, Z_2, .., T_i, Z_i)$. Notice that in the univariate case ($M = 1$) the definitions simplify to give $N(t) = \sum_{i\geq 1} 1(T_i \leq t)$, $\mathcal{F}_t^N = \sigma(N(s) : 0 \leq s \leq t)$ and $\mathcal{F}_{T_i}^N = \sigma(T_1, T_2, ..., T_i)$.

## 2.2 Stochastic Intensities and Likelihoods

As was stated earlier, a distinguishing feature of our approach is that the models are specified via the stochastic intensity.

**Definition 3 *Stochastic Intensity.*** *Let $N(t)$ be a simple point process on $[0, \infty)$ (defined on $(\Omega, \mathcal{F}, P)$) that is adapted to some filtration $\{\mathcal{F}_t\}$, and let $\lambda(t)$ be a positive, $\mathcal{F}_t$-predictable process. If*

$$E[N(t) - N(s)|\mathcal{F}_s] = E\left[\int_s^t \lambda(u)du|\mathcal{F}_s\right] \quad P\text{-}a.s., \tag{3}$$

*for all $s, t$ such that $0 \leq s \leq t$, then we say that $\lambda(t)$ is the $(P, \mathcal{F}_t)$-intensity of $N(t)$.*[10]

In words, (3) says that if one stands at time $s$ having observed the history $\mathcal{F}_s$ and wishes to predict the number of additional events that will occur by time $t$ using the conditional expectation of $(N_t - N_s)|\mathcal{F}_s$, then one can equivalently use $\mathrm{E}[\int_s^t \lambda(u)du|\mathcal{F}_s]$.[11] In this paper we shall usually be concerned with the case $\mathcal{F}_t = \mathcal{F}_t^N$. However, a strength of an intensity based framework is that it is straightforward conceptually to model the evolution of the PP conditional on a wider filtration, $\mathcal{F}_t$ (where $\mathcal{F}_t \supset \mathcal{F}_t^N \; \forall t$) that could include, for example, the internal history of some covariate continuous time process.

---

[8]The $j$th event, with occurrence time $T_j(\omega)$, is said to be $m$-type (for that particular $\omega$) iff $Z_j(\omega) = m$.

[9]Although we refer to the internal history of the process, the term natural filtration is widely encountered and would also be a suitable description here.

[10]Note that: 1) sufficient conditions for $\lambda(t)$ to be $\mathcal{F}_t$-predictable are that the sample paths of the process are left-continuous and have right-hand limits, and that $\lambda(t)$ is adapted to $\{\mathcal{F}_t\}$; 2) the conditional expectations in (3) are those defined with respect to the probability measure $P$.

[11]Under the conditions of definition (3) and the additional assumption that the point process $N(t)$ is integrable (that is, $\mathrm{E}[N(t)] < \infty$, $t \geq 0$), (3) is equivalent to the statement that $N(t) - \int_0^t \lambda(u)du$ is a $(P, \mathcal{F}_t)$-martingale. That is, the difference between the PP and its compensator is a martingale.

In the multivariate case ($M > 1$), the $M$-variate process $\lambda(t) = (\lambda_m(t))_{m=1}^M$ is said to be the $(P, \mathcal{F}_t)$-intensity of $N(t) = (N_m(t))_{m=1}^M$ if and only if $\lambda_m(t)$ is the $(P, \mathcal{F}_t)$-intensity of the counting process $N_m(t)$ for $m = 1, ..., M$.[12] As will be made clear later, a multivariate PP model can thus be determined by the specification of the $M$ intensities of the (univariate) counting processes $N_m(t)$. This perspective is very natural and productive if one is concerned with developing new models – one considers how the intensity for $m$-type events changes in continuous time as new 'information' arrives. The apparent problem with non $m$-type events occurring during a duration between $m$-type events causes no difficulty whatsoever in this framework since the path of the intensity can be made to depend on such events. Thinking in terms of an intensity for $m$-type events is made more intuitive by the fact that the intensity can often be interpreted as a conditional hazard – that is, as the conditional expectation in the limit of the number of $m$-type events that will occur per unit time in the future, given the current information set. We now make this connection explicit.

Consider a simple PP on $[0, T]$, $\{N(t)\}_{0 \le t \le T}$, with the $(P, \mathcal{F}_t)$-intensity $\lambda(t)$, where $\int_0^t \lambda(s)ds < \infty$ P-a.s. for all $t$ (which implies that, $P$-a.s., $N_t < \infty$ for all $t$) and the sample paths of $\lambda(t)$ are left-continuous with right-hand limits. Under the additional condition that $\{\lambda(t)\}_{0 \le t \le T}$ is bounded by an integrable random variable (hereafter r.v.), it is possible to show (see Lemma 3.3, Aalen (1978)) that

$$\lim_{s \downarrow t} \frac{1}{s - t} \mathrm{E}[N(s) - N(t)|\mathcal{F}_t] = \lambda(t+) \quad P\text{-a.s.}, \tag{4}$$

where $\lambda(t+) = \lim_{s \downarrow t} \lambda(s)$.

The following theorem is of fundamental importance since it establishes that legitimate likelihood functions and ratios are available for our statistical models, and expresses the likelihood in terms of the stochastic intensity.

**Theorem 2.1** *Let $0 < T < \infty$ and $(N_m(t))_{m=1}^M$ be an $M$-variate point process on $[0, T]$, defined on $(\Omega, \mathcal{F})$ (where $\mathcal{F}_T^N \subseteq \mathcal{F}$), such that $N_m(t)$ has $(P_0, \mathcal{F}_t^N)$-intensity equal to 1 for $m = 1, ..., M$. Suppose that $\lambda(t) = (\lambda_m(t))_{m=1}^M$ is an $M$-variate, positive, $\mathcal{F}_t^N$-predictable process satisfying*

$$\sum_{m=1}^M \int_0^T \lambda_m(s)ds < \infty \quad P_0\text{-a.s.} \tag{5}$$

---

[12]Note that $N_m(t)$ is also a simple PP and so Definition 3 applies straightforwardly here to each $\lambda_m(t)$.

*Define the probability measure $P$ (satisfying $P \ll P_0$) by the following density with respect to $P_0$.*[13]

$$\frac{dP}{dP_0}\Big|_{\mathcal{F}_T^N} = \exp\left\{\sum_{m=1}^{M}\left[\int_0^T (1 - \lambda_m(s))ds + \int_{(0,T]} \log\lambda_m(s)dN_m(s)\right]\right\}. \qquad (6)$$

*Then $N_m(t)$ has $(P, \mathcal{F}_t^N)$-intensity $\lambda_m(t)$ (for $m = 1, ..., M$).*[14] *Let $\tilde{P}$ be another probability measure such that $N(t)$ also has $(\tilde{P}, \mathcal{F}_t^N)$-intensity $\lambda(t)$. Then, $\tilde{P}(A) = P(A)$ for all $A \in \mathcal{F}_T^N$.*

**Proof.** Repeat the proof of Theorem 5.2, Karr (1991) setting the baseline intensity equal to $(1, 1, ..., 1) = \mathbf{1}_M$. For the uniqueness of $P$, see Brémaud (1981), Theorem T8, p.64. ∎

## 2.3 Intensity-based inference

### 2.3.1 Statistical models

In this paper each model is specified via a parametric family of stochastic intensities $\{\lambda_\theta(t)\}_{\theta \in \Theta}$, where $\lambda_\theta(t)$ is an $M$-variate, positive, $\mathcal{F}_t^N$-predictable process satisfying (5) for all $\theta \in \Theta$. Theorem 2.1 establishes the existence of an $M$-variate PP $N(t)$ and a unique probability measure $P_\theta$ such that $N(t)$ has $(P_\theta, \mathcal{F}_t^N)$-intensity $\lambda_\theta(t)$.. The implication is that we need only specify the parametric family of stochastic intensities in order to specify the statistical model completely. In econometrics, the family $\{P_\theta\}_{\theta \in \Theta}$ would often be referred to as the potential Data Generating Processes (DGPs). Since $\{P_\theta\}_{\theta \in \Theta}$ is dominated by $P_0$ (the law of a multivariate Poisson process), equation (6) constitutes a legitimate likelihood function. An advantage of the models that we propose is that, in all cases, integration of the sample path of the intensity can be performed analytically. This is extremely convenient when performing likelihood-based inference – since the intensity appears in the integrand of the first integral in (6) – and computing the diagnostic tests that we propose below.

### 2.3.2 Asymptotic distribution theory

Ogata (1978) establishes conditions under which the MLE for a simple, stationary (univariate) PP model is consistent and asymptotically normal as $T \to \infty$, and under which the LR test of a

---

[13]Note that evaluating the second integral in (6) yields $\sum_{i:0<T_i^{(m)}\leq T} \log\lambda_m(T_i^{(m)})$.

[14]The Poisson processes $N_1(t), ..., N_M(t)$ are independent under $P_0$. Since Poisson processes are $\tilde{P}$-independent iff they have no jumps in common $\tilde{P}$-a.s., it follows from $P \ll P_0$ that $N(t)$ has no common jumps $P$-a.s. That is, the pooled process $\sum_{m=1}^{M} N_m(t)$ is simple $P$-a.s.

simple null hypothesis possesses the standard $\chi^2$ asymptotic null distribution. Since the focus is on stationary PPs, the model is taken to be for a PP on $(-\infty, +\infty)$, $\{T_i; i = 0, \pm 1, \pm 2, ...\}$, which is observed during the interval $[0, T]$.[15] Let $\lambda^*(t)$ denote the complete intensity – that is the $\mathcal{H}_{(-\infty, t]}$-intensity, where $\mathcal{H}_{(-\infty, t]}$ is the 'infinite history' of the process – and (for the purposes of Section 2.3.2 only) let $\lambda(t) = \mathrm{E}[\lambda^*(t)|\mathcal{H}_{(0, t]}]$.[16] The exact log-likelihood for the statistical model, $l_T(\theta)$, is then given by the logarithm of (6) with $\lambda(t) = \mathrm{E}[\lambda^*(t)|\mathcal{H}_{(0, t]}]$ (and $M = 1$). Define also the (theoretical) conditional log-likelihood under the information from the infinite past, $l_T^*(\theta)$, by replacing $\lambda(t)$ with $\lambda^*(t)$ in the expression for $l_T(\theta)$. The central results of Ogata (1978) are summarised in the following theorem.

**Theorem 2.2** *Let the true parameter value be denoted by $\theta_0 \in \mathcal{R}^d$, and suppose that the regularity conditions A, B and C of Ogata (1978) are satisfied. Then, the MLE, $\hat{\theta}_T$, converges in probability to $\theta_0$ as $T \to \infty$. Suppose further that $\theta_0$ is in the interior of the parameter space. Then, as $T \to \infty$*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \to \mathcal{N}(0, I(\theta_0)^{-1}), \tag{7}$$

*in law, where $I(\theta_0)$ is the matrix $\{I_{ij}(\theta_0)\}_{ij}$ and*

$$I_{ij}(\theta_0) = E_{\theta_0} \left[ \frac{1}{\lambda_{\theta_0}^*(t)} \frac{\partial \lambda_{\theta_0}^*(t)}{\partial \theta_i} \frac{\partial \lambda_{\theta_0}^*(t)}{\partial \theta_j} \right] = -E_{\theta_0} \left[ \frac{1}{T} \frac{\partial^2 l_T^*(\theta_0)}{\partial \theta_i \partial \theta_j} \right]. \tag{8}$$

*Under the above assumptions,*

$$E_{\theta_0} \left\{ \int_0^T \left[ \frac{1}{\lambda_{\theta_0}(s)} \frac{\partial \lambda_{\theta_0}(s)}{\partial \theta_i} \frac{\partial \lambda_{\theta_0}(s)}{\partial \theta_j} \right] ds \right\} = -E_{\theta_0} \left[ \frac{\partial^2 l_T(\theta_0)}{\partial \theta_i \partial \theta_j} \right] \quad \forall i, j, \tag{9}$$

*and*

$$-E_{\theta_0} \left[ (1/T)(\partial^2 l_T(\theta_0)/\partial \theta_i \partial \theta_j) \right] \to I_{ij}(\theta_0) \quad \forall i, j. \tag{10}$$

*Note that (9) is also valid when $\{\lambda_{\theta_0}, l_T(\theta_0)\}$ is replaced by $\{\lambda_{\theta_0}^*, l_T^*(\theta_0)\}$. Also, under the above assumptions,*

$$2[l_T(\hat{\theta}) - l_T(\theta_0)] \to \chi_d^2, \tag{11}$$

*in law.*

---

[15]Note that this is in contrast to the models used in this paper in which the counting processes start at $t = 0$.

[16]We define $\mathcal{H}_{(s, t]}$ to be the sigma field $\sigma(N(t) - N(u); s < u \leq t)$. For a rigorous definition of a simple PP on $(-\infty, +\infty)$ and of an $\mathcal{H}_{(-\infty, t]}$-stochastic intensity for such a process, see Brémaud (1996). These definitions are the natural analogues of the definitions we have given for a simple PP on $[0, +\infty)$.

***Proof.*** *See Theorem 1 and its subsequent remark, Theorem 2 and Theorem 5 of Ogata* *(1978).* ∎

These are essentially the standard asymptotic properties of a MLE, but with an additional subtlety introduced: it is the complete intensity, $\lambda_\theta^*(t)$, rather than $\lambda_\theta(t)$ that appears in (8) and the conditions of the theorem do not imply that $I_{ij}(\theta_0) = -\mathrm{E}_{\theta_0}\left[(1/T)(\partial^2 l_T(\theta_0)/\partial\theta_i\partial\theta_j)\right]$. Nevertheless, the convergence result in (10) motivates the use of the normed Hessian matrix, $(1/T)(\partial^2 l_T(\hat\theta_T)/\partial\theta\partial\theta')$, to estimate $\{I_{ij}(\theta_0)\}_{ij}$.[17]

The development of further asymptotic distribution theory following on from the work of Ogata (1978) is an important area for future research. Results for the $M$-variate and non-stationary cases would be of particular interest in our context.

## 3    Diagnostic Testing

Consider the sequence $\{e_i^{(m)}\}_i$, obtained by integrating the sample path of the $\mathcal{F}_t^N$-intensity for $m$-type events over each $m$-type duration (that is, the time interval between adjacent $m$-type events). Russell (1999) suggests the use of this sequence as the basis for diagnostic testing and indicates that the results of Yashin and Arjas (1988) might be used to establish the *i.i.d.* exponential property of this sequence provided that a certain absolute continuity condition is satisfied. In our discussion of Russell's paper in Section 6.2 we give reasons why this particular condition is very restrictive in the context of multivariate PP models for market event data, thus highlighting the need for a general proof that relies on much weaker assumptions concerning the PP model. We show in Section 3.1 how the *i.i.d.* exponential property of the $\{e_i^{(m)}\}_i$ sequence can be derived by a random change of time argument. It turns out that the condition that the stochastic intensity must satisfy in order for this proof to apply is extremely natural in the context of models of financial market events. We thus establish the broad applicability of diagnostic testing procedures based on the $\{e_i^{(m)}\}_i$ sequences ($m = 1, ..., M$) to multivariate PP models of financial market event data. In Section 3.2, various diagnostic tests are proposed, some of which are new to the econometrics PP literature, and the issue of how to combine the

---

[17]Ogata (1978) comments on the validity of the assumptions of Theorem 2.2 for the HawkesE(1) model defined later by equation (18) with $\mu(t) = \mu \; \forall t$.

information in the $M$ sequences is considered. A treatment of possible approaches to the impact of parameter uncertainty on the testing procedures is deferred until Section 6.3.1 in order to maintain the flow of the exposition of the main subject matter of the paper.

## 3.1 A random change of time argument

Identifying the $\{e_i^{(m)}\}_i$ sequence as the durations of a standard Poisson process obtained by transforming the PP, $N_m(t)$, using a random change of time allows us then to establish the $i.i.d.$ exponential property of the sequence immediately. Theorem 3.1 below (derived from Theorem T16, p.41 of Brémaud (1981)) forms the basis for the proof of the result in Corollary 3.2. The main contribution of this section is not the Corollary itself, which is no doubt well known to probabilists, but lies rather in establishing the connection between change of time techniques and diagnostic testing for parametric PP models. The benefit gained by introducing these techniques from the stochastic process literature is a rigorous proof of the $i.i.d.$ exponential property of the diagnostic sequence, $\{e_i^{(m)}\}$, suggested by Russell (1999) and hence the isolation of (12) as a sufficient condition for the validity of such diagnostic testing procedures. Random changes of time also appear to have been exploited surprisingly little to construct goodness of fit tests (for point and counting processes) in both the theoretical and applied statistical literature.[18] In the case where a single realisation of the counting process is observed, goodness of fit tests are sometimes based on the difference, $N(t) - \int_0^t \lambda(s)ds$, which is a mean zero martingale. An advantage of the change of time approach is the relative ease of establishing distributional results (at least in the absence of parameter uncertainty).

**Theorem 3.1** *Let $N(t)$ be an $M$-variate point process on $[0, \infty)$ ($M \geq 1$) with internal history $\mathcal{F}_t^N$ and let $N_m(t)$ ($m = 1, ..., M$) be the associated 'marginal' counting processes as given by Definition 2. Let $\mathcal{F}_t$ be a history of $N(t)$ (that is, $\mathcal{F}_t^N \subseteq \mathcal{F}_t \ \forall t \geq 0$), and suppose, for each $m$, that $N_m(t)$ has the $(P, \mathcal{F}_t)$-intensity $\lambda_m(t)$ where $\lambda_m(t)$ satisfies*

$$\int_0^\infty \lambda_m(t)dt = \infty \quad P\text{-}a.s. \tag{12}$$

---

[18]Exceptions of which we are aware are Arjas (1986) and Arjas and Haara (1988). They exploit the exponentiality of the integrated hazards (intensities) and their independence across individuals in failure time models. The hypothesis test in the empirical application of Aalen and Hoem (1978) could also be viewed as a sort of goodness of fit test based on a random change of time.

*Define for each $m$ and all $t$, the $\mathcal{F}_t$-stopping time $\tau_m(t)$ as the (unique) solution to*[19]

$$\int_0^{\tau_m(t)} \lambda_m(s)ds = t. \tag{13}$$

*Then for each $m$, the point process $\tilde{N}_m(t)$ defined by*

$$\tilde{N}_m(t) = N_m(\tau_m(t)), \tag{14}$$

*is a standard Poisson process (that is, $\tilde{N}_m(t)$ is a $(P, \mathcal{F}_{m,t}^N)$-Poisson process with intensity 1, where $\mathcal{F}_{m,t}^N = \sigma(N_m(s) : 0 \le s \le t)$).*

**Proof.** Repeat the proof of Theorem T16, p.41, Brémaud (1981) replacing $N_t$ by $N_m(t)$ and setting (in the notation used there) $\mathcal{G}_t = \mathcal{F}_t$. ∎

Theorem T16 of Brémaud (1981) is stated for a univariate PP. Extending the theorem to the multivariate case where the change of time is applied to the 'marginal' counting processes, $N_m(t)$, of an $M$-variate PP (using the $\mathcal{F}_t$-intensity of each, $\lambda_m(t)$) is straightforward since $N_m(t)$ is itself a univariate PP.[20] Since we allow $\mathcal{F}_t^N \subset \mathcal{F}_t$ here, our diagnostic testing procedures can also be applied to models where we condition on a wider filtration than the internal history of the PP. We note that Aalen and Hoem (1978) prove additionally that $\tilde{N}_1(t), ..., \tilde{N}_M(t)$ are independent Poisson processes.[21]

**Corollary 3.2** *Let $\{T_i^{(m)}\}_{i \in \{1,2..\}}$ ($m = 1, ..., M$) be the sequences of points associated with the $N_m(t)$ in Theorem 3.1, and define for each $m$, $T_0^{(m)} := 0$, $e_i^{(m)} := \int_{T_i^{(m)}}^{T_{i+1}^{(m)}} \lambda_m(s)ds$ ($i = 0, 1, 2, ...$). Then $\{e_i^{(m)}\}_{i \in \{0,1,2..\}}$ is an i.i.d. sequence of Exponential random variables with mean 1 for $m = 1, ..., M$.*

**Proof.** Let $\{\tilde{T}_i^{(m)}\}_{i \in \{1,2..\}}$ be the sequence of points associated with $\tilde{N}_m(t)$ and define $\tilde{T}_0^{(m)} := 0$. Let $s$ satisfy $\tilde{T}_{i-1}^{(m)} \le s < \tilde{T}_i^{(m)}$ ( for some $i \ge 1$). Then (14) $\Rightarrow \tau_m(s) < T_i^{(m)} \le \tau_m(\tilde{T}_i^{(m)})$.

---

[19]Recall that $\lambda_m(s) > 0$ $\forall s$ from our definition of a stochastic intensity. This avoids problems with multiple solutions to (13) that could occur if we allowed $\lambda_m(s) = 0$.

[20]Theorem T16, p.41 of Brémaud (1981) does not require the intensity to be with respect to the *internal* history of the univariate PP in question ($\mathcal{F}_{m,t}^N$ in this case). Rather, the requirement is that the univariate PP be adapted to the history with respect to which the intensity is defined. This requirement is clearly satisfied here since $\mathcal{F}_{m,t}^N \subset \mathcal{F}_t^N \subseteq \mathcal{F}_t$ $\forall t$.

[21]See Section 4.5 of Aalen and Hoem (1978); they require that $\mathcal{F}_t = \mathcal{F}_t^N$ and that each 'distribution function' $\Pr[T_{i+1} - T_i \le x, Z_{i+1} = m | \mathcal{F}_{T_i}^N]$ is absolutely continuous in $x$ and has a derivative which is left continuous and has right-hand limits at each point. Importantly, Meyer (1971) establishes the result in the more general case where $\mathcal{F}_t^N \subset \mathcal{F}_t$ $\forall t$.

Since $\tau_m$ is continuous, letting $s \uparrow \tilde{T}_i^{(m)}$ establishes that $\tau_m(\tilde{T}_i^{(m)}) = T_i^{(m)}$. It then follows from (13) that $\tilde{T}_{i+1}^{(m)} - \tilde{T}_i^{(m)} = \int_{T_i^{(m)}}^{T_{i+1}^{(m)}} \lambda_m(s)ds$ for $i = 0, 1, 2, ...$ The result is then immediate since the $e_i^{(m)}$'s are the durations of a standard Poisson process. ■

Note that the only condition that we require is given in (12) and that this holds if and only if $\lim_{t\uparrow\infty} N_m(t) = \infty$ $P$-a.s. (see Lemma L17, p.41, Brémaud (1981)). This condition is very natural in the context of models of financial market events since it is equivalent to zero probability being assigned to sample paths in which no more $m$-type events ever occur after some point in time.

## 3.2  Constructing diagnostic tests

We now go on to consider how to use Corollary 3.2 in order to construct diagnostic tests. It is worth noting at this point that whilst it is often the case for univariate models that the $\{e_i\}$ series can be obtained by transforming the particular *i.i.d.* uniform series used for diagnostic evaluation by authors such as Shephard (1994), Kim, Shephard, and Chib (1998) and Diebold, Gunther, and Tay (1998) into exponential r.v.'s, our diagnostic procedures for multivariate PPs cannot be interpreted as an implementation of multivariate extensions of this approach such as that described in Diebold, Hahn, and Tay (1999). Some additional explanation of this point is given later in connection with the discussion of the impact of parameter uncertainty in Section 6.3.1.

Since the $M$-variate PP has the $(P_\theta, \mathcal{F}_t)$-stochastic intensity $\lambda_\theta(t) = (\lambda_m(t; \theta))_{m=1}^M$, Corollary 3.2 implies that if the DGP is $P_{\theta_0}$, then $e_i^{(m)}(\theta_0) \sim$ *i.i.d.* $Exp(1)$, where

$$e_i^{(m)}(\theta) = \int_{T_i^{(m)}}^{T_{i+1}^{(m)}} \lambda_m(s; \theta)ds, \tag{15}$$

for $m = 1, ..., M$. There are many possibilities for the construction of diagnostic tests based on the $M$ $\{e_i^{(m)}(\theta_0)\}_i$ series that spring to mind.[22] Consider first tests that are based say the $m$th individual series. Tests that use the series directly include Box-Ljung tests of zero autocorrelation of the series itself and the squared series, and tests based on the moments of the exponential distribution. One such is the test of no excess dispersion suggested by Engle and

---

[22]In practice, the aim is to test whether there exists a $\theta_0$ such that the DGP is given by $P_{\theta_0}$. The impact of the attendant parameter uncertainty (or 'nuisance parameter' problem) is discussed in Section 6.3.1.

Russell (1998) which is given by $\sqrt{N_m(T)}((\hat{\sigma}^2_{e(m)} - 1)/\sqrt{8})$, where $\hat{\sigma}^2_{e(m)}$ is the sample variance of the $e_i^{(m)}(\theta_0)$. The test statistic has an asymptotic $\mathcal{N}(0, 1)$ distribution under the null hypothesis that $\{e_i^{(m)}(\theta_0)\}$ is $i.i.d.$ $Exp(1)$.[23] Alternatively, the series can be transformed to yield an $i.i.d.$ uniform series under the null. Thompson (2001) derives the asymptotic distribution of various test statistics based on the centred empirical distribution function and the normalised cumulative periodogram under the null that the series is $i.i.d.$ uniform and explains how to compute quantiles of the (finite sample) null distributions by simulation.[24]

Other diagnostic testing possibilities arise by combining the $M$ series in some way, either prior to or after the calculation of the test statistics. We noted above that Aalen and Hoem (1978) have proved under quite general conditions that, under the null of correct specification, the transformed processes, $\tilde{N}_1(t), ..., \tilde{N}_M(t)$, are independent Poisson processes each with intensity 1. It follows that the superposition $\sum_{m=1}^{M} \tilde{N}_m(t)$ is a Poisson process with intensity $M$. Tests of the independence property of the transformed processes may therefore be constructed by 'pooling' the points associated with the $M$ $\{e_i^{(m)}(\theta_0)\}$ series (i.e. sorting the collection of points $\{\tilde{T}_i^{(m)}\}_{i,m}$ into ascending order) and then calculating tests of the $i.i.d.$ $Exp(1)$ property of the durations between those points (after rescaling the durations by multiplying by $M$) as described above.[25] We have not pursued the possibility of either combining multiple tests based on the $m$th series into a single test or combining tests based on different series here. The distributional results of Thompson (2001) can be used to achieve the former. Two tests based on, say, the $m$th and $q$th ($m \neq q$) series will be independent under the conditions of Aalen and Hoem (1978).

In the empirical sections of the paper, we routinely report 3 diagnostic test statistics for each $m = 1, ..., M$ : the Box-Ljung tests that the first 15 autocorrelations are all equal to zero for the $\{e_i^{(m)}(\hat{\theta})\}$ and $\{(e_i^{(m)}(\hat{\theta}))^2\}$ series, and the test for excess dispersion based on the $\{e_i^{(m)}(\hat{\theta})\}$ series. $\hat{\theta}$ here denotes the MLE. The corresponding $p$-values are calculated using the asymptotic

---

[23]Note that 8 is the variance of $\left(e_i^{(m)}(\theta_0) - 1\right)^2$ under the null hypothesis.

[24]The evaluation of model specification using a transformation of the data that is approximately $i.i.d.$ uniform has a long history, dating back to Rosenblatt (1952). Smith (1985) and Shephard (1994) were among the first to use the 'integral transform' method to evaluate time series models.

[25]In practice $N_1(t), ..., N_M(t)$ are observed on $[0, T]$, $T < \infty$, but $\tilde{N}_m(t)$ and $\tilde{N}_q(t)$ ($m \neq q$) are usually observed over different intervals as a result of the change of time. We adopt the following pragmatic approach to this problem in Section 5.4. Let $C_m$ be the last observed point of $\tilde{N}_m(t)$ and $C := \min\{C_m, m = 1, ..., M\}$. Then we sort $\{\tilde{T}_i^{(m)} : \tilde{T}_i^{(m)} \leq C, m = 1, ..., M\}$ into ascending order and then proceed as before.

null distributions which hold in the case of known parameters. The issue of how to take account of the impact of parameter uncertainty on the testing procedures is taken up in Section 6.3.1.[26] In addition, when $M > 1$, the same 3 statistics are also reported using the series of durations obtained by pooling the points associated with the $M$ $\{e_i^{(m)}(\hat{\theta})\}$ series (see above).

# 4   Univariate intensity-based models

Our aim in this paper is to develop a continuous time econometric modelling framework that is applicable to multivariate market event data. Nonetheless, we have chosen to describe first a model for univariate PPs. This allows the main ideas to be presented in a relatively simple setting so that extending the discussion to the bivariate case in Section 5 will then be more straightforward as the univariate and bivariate models share exactly the same basic 'building blocks.'

Section 4.2 defines our univariate generalised Hawkes process, elucidates its relationship to the motivating self-exciting process of Hawkes (1971), and discusses hypothesis testing and the computation of MLEs. An empirical application of the model is then presented in Section 4.3 for the timing of trades in two different corporate stocks, one listed on the NASDAQ (National Association of Securities Dealers Automated Quotation) system and the other on the NYSE. In order to clarify the presentation that follows, we first discuss the data transformation that is used throughout the paper.

## 4.1   Data transformation

The empirical applications of our models will consider the timing of trades and the timing of certain changes to the mid-quote that occur during normal trading hours – that is, during the 6.5 hour period between 9:30 EST and 16:00 EST. Since the equity markets in question do not operate continuously, the researcher is faced with the question of how to model data that is generated during trading days that are separated by overnight periods when the market is closed. We take the following approach throughout the paper. Time zero is taken to be 9:30 on the first trading day of the sample and the data pertaining to each of the 6.5 hour trading

---

[26]The approach to the parameter uncertainty adopted in the empirical sections of the paper is in line with most previous work in the econometrics literature.

days is then concatenated in order to remove the overnight periods. Thus, the occurrence times of the market events are mapped onto $[0, \infty)$ as follows: if $x$ is the time in hours after 9:30 of an event occurring on the $d$th trading day included in the dataset ($d = 1, 2, ...$), then that event appears as an event at time $x + 6.5(d - 1)$ in our final dataset.[27]

Two comments concerning the above transformation are in order. First, transforming the data in this way and treating it as a realisation of a single PP on $[0, T]$ allows us to make straightforward use of existing theorems in the PP literature. The alternative would be to view the data for each trading day as the realisation of a different PP and to specify the nature of the dependence between these PPs. This is unnecessarily complicated. One must be able to specify how the intensity on a particular trading day depends on the entire history of events (including those of previous trading days), but this can be achieved within the 'single point process' framework adopted here. However, it is important then to take into account the overnight periods during which the market is closed and to model carefully how the intensity today depends on events that occurred on previous trading days in real time. This is the motivation for the generalised Hawkes models introduced below. Second, other features such as events occurring outside normal trading hours and stock exchange opening procedures are excluded from the present analysis. This reflects the particular focus of our empirical work which is the modelling of the operation of the continuous market during normal trading hours. Nevertheless, it is in principle straightforward within our modelling framework to condition on such information by allowing the intensity at the start of the trading day to depend upon these additional events (see Section 6.3.2).[28]

## 4.2 Generalised Hawkes Processes

### 4.2.1 Representations of the g-HawkesE(k) process

The intensity of the generalised Hawkes (g-Hawkes) process is defined recursively in terms of the levels of the non-deterministic components of the intensity at the end of the $(d - 1)$th trading day and the contributions of the events occurring on day $d$. In accordance with the

---

[27] Note that the units of measurement for 'wall-clock' time, $t$, are hours in the empirical sections of this paper.

[28] We have not developed such an approach here in any detail. A preliminary analysis for NASDAQ biotechnology stocks of quoting activity in the preopening period (including the occurrence of 'crossed and locked' quotes) failed to reveal a significant effect on the intensity of quote changes during normal trading hours.

data transformation that we employ, we partition the real half-line into intervals of length $l$ that correspond to the different trading days. This partition is written as

$$(0, \infty) = (0, \tau_1] \cup (\tau_1, \tau_2] \cup ... \cup (\tau_{d-1}, \tau_d] \cup ...,$$

where $\tau_d = l \cdot d$ $(d = 0, 1, 2, ...)$. $\tau_d$ denotes the time corresponding to the end of the $d$th trading day.[29]

**Definition 4 *Univariate g-HawkesE(k) model.*** *The model is defined by the stochastic intensity*

$$\lambda(t) = \mu(t) + \sum_{j=1}^{k} \tilde{\lambda}_j(t), \tag{16}$$

*where $\mu(t)$ is a positive, deterministic function of time and, for $j = 1, 2, ..., k$,*

$$\tilde{\lambda}_j(t) = \pi_j \tilde{\lambda}_j(\tau_{d-1}) e^{-\rho_j(t-\tau_{d-1})} + \int_{[\tau_{d-1}, t)} \alpha_j e^{-\beta_j(t-u)} dN(u), \tag{17}$$

*for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ....)$, and $\tilde{\lambda}_j(0) = 0, \alpha_j \geq 0, \beta_j > 0, \pi_j \geq 0, \rho_j > 0$.[30]*

The stochastic intensity of the g-HawkesE($k$) model is thus the sum of a deterministic component, $\mu(t)$, and $k$ non-deterministic components, $(\tilde{\lambda}_j(t))_{j=1}^{k}$. Equation (17) expresses each $\tilde{\lambda}_j(t)$ as the sum of the exponentially-damped value of $\pi_j \cdot \tilde{\lambda}_j(\tau_{d-1})$, where $\tilde{\lambda}_j(\tau_{d-1})$ is the level of the $j$th component at the end of the previous trading day, and the contributions of events occurring on day $d$ and prior to time $t$. Thus $\tilde{\lambda}_j(t)$ depends on the timing of events occurring prior to day $d$ only via the term $\tilde{\lambda}_j(\tau_{d-1})$. We call the first term in (17) the ($j$th) intensity 'spillover effect' between trading days. Evaluating the second term yields $\sum_{i:\tau_{d-1} \leq T_i < t} \alpha_j e^{-\beta_j(t-T_i)}$. We refer to this as the ($j$th) 'Hawkes-type' term for reasons that will be made explicit later. A sample path of this term is left-continuous, jumping up by an amount equal to $\alpha_j$ in response to the occurrence of an event and then decaying exponentially (according to the parameter $\beta_j$) until the occurrence of the next event.

It is clear that the g-HawkesE($k$) model allows the dependence of the intensity on events occurring on previous trading days to be estimated from the data. The model nests the important

---

[29]$l = 6.5$ in all of the empirical sections of the paper.

[30]The 'E($k$)' refers to the superposition of $k$ non-deterministic components, $\tilde{\lambda}_j(t)$, each of which is specified in terms of an Exponential 'response function,' $\alpha_j e^{-\beta_j s}$. This function gives the current 'response' to an event that occurred $s$ time units ago. Clearly, other specifications of the response functions are possible, but are not considered here.

case where there is no dependence between trading days. This occurs when $\pi_j = 0 \; \forall j$. The g-HawkesE($k$) process is also equivalent to the 'self-exciting' process of Hawkes (1971) when the restrictions $\pi_j = 1, \rho_j = \beta_j$ for $j = 1, ..., k$ are imposed.[31] Note that this imposes a very restrictive form on the spillover effects. The Hawkes (1971) model – referred to here as the HawkesE($k$) model – is given by

$$
\begin{aligned}
\lambda(t) &= \mu(t) + \int\limits_{(0,t)} \sum_{j=1}^{k} \alpha_j e^{-\beta_j(t-u)} dN(u) \qquad (18) \\
&= \mu(t) + \sum_{j=1}^{k} \sum_{i:0<T_i<t} \alpha_j e^{-\beta_j(t-T_i)},
\end{aligned}
$$

where $\mu(t)$ is again a positive, deterministic function of time and $\alpha_j \geq 0, \beta_j > 0$ for $j = 1, ..., k$.[32] The second term in (17) is the same as the inner sum in (18) except that the sum is not taken over $T_i < \tau_{d-1}$. The equivalence of the g-HawkesE($k$) and HawkesE($k$) models when $\pi_j = 1, \rho_j = \beta_j \; \forall j$ is a corollary of the following proposition that establishes a representation of the g-HawkesE($k$) model in terms of the contributions of the events occurring on the previous trading days.

**Proposition 4.1** *Define $\breve{\lambda}_j(t)$ for $j = 1, ..., k$ by*

$$
\breve{\lambda}_j(t) = \sum_{s=1}^{d-1} \pi_j^s C_j(d-s) e^{-\rho_j(t-l(d-s))} + \int\limits_{[\tau_{d-1},t)} \alpha_j e^{-\beta_j(t-u)} dN(u), \qquad (19)
$$

*for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ....)$ and $\breve{\lambda}_j(0) = 0$, where*

$$
C_j(d) = \int\limits_{[\tau_{d-1},\tau_d)} \alpha_j e^{-\beta_j(\tau_d-u)} dN(u), \qquad (20)
$$

*is the contribution of the events occurring on day $d$. Then if $\breve{\lambda}(t) = \mu(t) + \sum_{j=1}^{k} \breve{\lambda}_j(t)$, and $\lambda(t)$ is the g-HawkesE($k$) model defined above by (16) and (17), then the processes $\lambda(t)$ and $\breve{\lambda}(t)$ are indistinguishable – that is $\lambda(t)_\omega = \breve{\lambda}(t)_\omega, t \geq 0$ for all $\omega \in \Omega$.*

    **Proof.** *Since the proof is straightforward only the basic approach is indicated here. Fix an $\omega \in \Omega$ and hence $\{T_i(\omega)\}$. Clearly $\lambda(0)_\omega = \breve{\lambda}(0)_\omega = \mu(0)$. Then show that $\lambda(t)_\omega = \breve{\lambda}(t)_\omega \; \forall t > 0$*

---

[31]Despite their usefulness, Hawkes-type models do not seem to have been used before in financial econometrics. The probabilistic properties of Hawkes processes are discussed in Hawkes (1971), Hawkes and Oakes (1974), and Brémaud (1996); ML estimation of self-exciting models is considered by Ogata (1978), and Ozaki (1979), Ogata and Akaike (1982); and application of such models to earthquake data is treated by Vere-Jones and Ozaki (1982) and Ogata (1983).

[32]Strictly speaking, the original Hawkes (1971) model is given by (18) with $\mu(t) = \mu \; \forall t$.

*by establishing that, for $j = 1, .., k$ and $d = 1, 2, ...,$ $\tilde{\lambda}_j(t)_\omega = \breve{\lambda}_j(t)_\omega, \tau_{d-1} < t \leq \tau_d$. The latter fact may be established by induction on $d$.* ∎

**Corollary 4.2** *The g-HawkesE(k) model with the parameter restrictions $(\pi_j = 1, \rho_j = \beta_j)$ for $j = 1, ..., k$ and the HawkesE(k) model in (18) are identical models (that is, the processes are indistinguishable). This follows easily by substitution into (19) and some straightforward algebraic manipulation.*

The representation of the g-HawkesE($k$) model given by Proposition 4.1 is intuitively appealing as it expresses the $j$th non-deterministic component of the intensity at time $t$ (on day $d$) as the summation of a Hawkes-type term that depends only on day $d$ events and the weighted sum of the contributions, $C_j(d - s)$, of previous trading days. Provided that $\pi_j > 0$, the weights are declining in $s$ iff $6.5\rho > \log \pi$ (a condition that is found to hold in all of our applications).[33]

The strength of our approach is that rather than imposing strong, *a priori* assumptions about the nature of the dependence between trading days, we are able to estimate the effect from the data. Our empirical results will illustrate the value of avoiding such assumptions in improving the goodness of fit of models for market event data. Adopting an inappropriate *a priori* assumption (such as a lack of dependence between trading days for some NASDAQ stocks) can result in the rejection of a model that is essentially well specified (apart, that is, from its treatment of the spillover effects). Similarly, fitting a model such as the HawkesE($k$) model (which might be more appropriate if trading were continuous) usually results in rejection of the model because the spillover effects it implies are too restrictive. Since the nature of the dependence is likely to vary across different exchanges, stocks and calendar periods it is clearly desirable to estimate the effect from the data.

Our approach thus recognises an important feature of financial markets: namely, that for the majority of markets, the market does not operate continuously. Previous work has tended not to focus on this feature of the data. Often a lack of dependence is assumed so that the

---

[33]It is also interesting to consider the contribution of the term that depends only on day $d$ events to the $j$th component, $\tilde{\lambda}_j(t)$, as $t$ increases (for $6.5d < t$). This component is of the form $K(t)C_j(d)$. When $\pi_j > 1$, $K(t)$ is strictly decreasing during the day but jumps up at the start of each day; when $\pi_j = 1$, $K(t)$ is continuous and strictly decreasing in $t$; and when $0 < \pi_j < 1$, $K(t)$ is again strictly decreasing during the day and jumps down at the start of each day.

intensity today depends only on the timing of trades that have occurred since today's opening (see, for example, Engle and Russell (1998)). Russell (1999) maps the data onto the real half line as we do and then treats it as the realisation of a (multivariate) PP on $[0, \infty)$. However, the approach adopted there is analogous to the HawkesE($k$) model in (18) in the following way: when there is an 'infinite past', identical past realisations of the process prior to time $t$ imply that the non-deterministic component of the intensity at time $t$ is the same whatever wall-clock time in the trading day $t$ corresponds to.[34]

### 4.2.2 Hypothesis testing and computation of MLEs

Consider testing the null $H_0 : \alpha_2 = 0$, against the alternative $H_1 : \alpha_2 > 0$ (the maintained hypothesis being that $\alpha_2 \geq 0$). This testing problem violates two of the regularity conditions that are assumed to hold in standard testing problems. The parameter value under the null lies on the boundary of the maintained hypothesis, and there are nuisance parameters $(\pi_2, \rho_2, \beta_2)$ that are identified under the alternative but not under the null. The same comments apply to a test of $H_0 : \pi_j = 0$ against the alternative $H_1 : \pi_j > 0$ (in which case $\rho_j$ is unidentified under the null). A consequence is that the LR tests of the various hypotheses will not possess the standard $\chi^2$ asymptotic null distributions. Such a situation is not uncommon in econometrics – consider, for example, a test of the null of no conditional heteroskedasticity in a GARCH(1,1) model – and is exactly the situation considered by Andrews (2001) in the context of time series models. Establishing analogous results for PP models is not a trivial task and is beyond the scope of the present paper. This matter will be explored in future work.

We now briefly describe the computation of MLEs for the univariate models.[35] In order to simplify the presentation, we assume that the process is observed over an integer number of trading days. Taking logarithms of (6) (with $M = 1$) and decomposing the resultant log-likelihood into the contributions of the different trading days yields

$$l(\theta) = \sum_{d=1}^{T/l} \left\{ \int_{A_d} (1 - \lambda_\theta(s)) ds + \int_{A_d} \log \lambda_\theta(s) dN(s) \right\}, \tag{21}$$

---

[34]By 'identical past realisations' we mean that the distance of all of the points from $t$ is the same in both cases.

[35]Note that when we view the g-HawkesE($k$) model as a statistical model we impose the additional constraints $\beta_1 > \beta_2 > ... > \beta_k$ so that it is identified.

where $A_d = (\tau_{d-1}, \tau_d]$ and $\lambda_\theta(s)$ is the intensity of the g-HawkesE($k$) model given by (16) and (17). Decomposing the log-likelihood in this way allows us to use the recursive specification given by (17) in order to compute the log-likelihood. Evaluating (21) yields[36]

$$
\begin{aligned}
l(\theta; T_1, ..., T_{N(T)}) \;=\; & T - \int_0^T \mu_\gamma(s)ds + \sum_{d=1}^{T/l} \sum_{T_i \in A_d} \log \lambda_\theta(T_i) - \\
& \sum_{d=1}^{T/l} \sum_{j=1}^k \left\{ \pi_j/\rho_j (1 - e^{-l\rho_j}) \tilde\lambda_{j,\theta}(\tau_{d-1}) + \sum_{\tau_{d-1} \le T_i < \tau_d} \alpha_j/\beta_j (1 - e^{-\beta_j(\tau_d - T_i)}) \right\}.
\end{aligned}
\tag{22}
$$

Numerical optimisation of $l(\theta; T_1, ..., T_{N(T)}) - T$ was performed using the BFGS algorithm with numerical derivatives (see Doornik (2001)). In order to estimate the g-HawkesE(2) model, initial parameter values for the algorithm were obtained by a sequential estimation procedure in which the estimates obtained from each model determine the starting values for the subsequent one. Full details of this procedure, the recursions used to improve computational efficiency and the reparametrisation of the log-likelihood employed to impose the parameter constraints of the model are described in the Appendix.

## 4.3   An application to the timing of trades on NASDAQ and the NYSE

An illustration of the empirical usefulness and flexibility of the univariate g-HawkesE($k$) model is provided here by an analysis of two different datasets obtained from the TAQ database: the timing of trades in the biotechnology stock Genzyme General (GENZ) on NASDAQ during the period from 1 March 2000 to 26 April 2000 and the timing of trades in The Walt Disney Company (DIS) on the NYSE from 5 July 2000 to 29 August 2000. Each of these datasets comprises 40 trading days. We model the timing of trades that occur during normal trading hours (i.e. 9:30 EST to 16:00 EST) only, applying the data transformation described in Section 4.1. Further details regarding amendments that we make to the TAQ data are provided in the Appendix.

Summary statistics of the durations between trades in the final datasets are given in Table 1.[37]  The mean intertrade duration for GENZ (13 seconds) is considerably shorter than that

---

[36]The specification and parametrisation of the deterministic component of the intensity, $\mu_\gamma(t) = \mu(t; \gamma_1, ..., \gamma_8)$, is discussed in the following section.

[37]The following practice is adopted for the reporting of test statistics throughout the paper: $p$-values are shown in parentheses and tests that reject at the 1% level are shown in bold.

|  | GENZ − 1.3.2000 to 26.4.2000 | DIS - 5.7.2000 to 29.8.2000 |
|---|---|---|
| No. of durations | 70,102 | 46,210 |
| Mean duration | 0.222 | 0.338 |
| Std deviation | 0.365 | 0.368 |
| Minimum | 1/60 | 1/60 |
| Maximum | 7.550 | 11.933 |
| $BL$ | **16 945** (0.000) | **2721.5** (0.000) |

Table 1: Summary Statistics of Intertrade durations for Genzyme General (GENZ) and Walt Disney Company (DIS). The summary statistics are for the series obtained by calculating the time intervals (durations) between trades. BL is the Box-Ljung statistic for zero autocorrelation calculated using the first 15 lags; the durations are measured in minutes.

for DIS (20 seconds). The autocorrelation functions of the durations (not shown here) exhibit long sequences of positive autocorrelations similar to those reported elsewhere for intertrade durations (see, for example, Engle and Russell (1998)).

It is well known that intradaily seasonality is an important feature of such data. In common with Russell (1999), we have found that adopting a piecewise linear spline (continuous over the course of the 6.5 hour trading day and with 8 nodes at 9:30,10:00,11:00,...,16:00) for the deterministic component $\mu_\gamma(t)$ works well in practice. Specifically, we set

$$\mu_\gamma(t) = \begin{cases} 1_{v(t)\in(0,0.5]}[\gamma_1 + 2v(t)(\gamma_2 - \gamma_1)] + \\ \sum_{i=1}^{6} 1_{v(t)\in(i-0.5,i+0.5]}[\gamma_{i+1} + (v(t) - i + 0.5)(\gamma_{i+2} - \gamma_{i+1})] \text{ for } v(t) > 0, \\ \gamma_8 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } v(t) = 0, \end{cases} \quad (23)$$

where $v(t) = 6.5(t/6.5 - \lfloor t/6.5 \rfloor)$ is the number of hours that have elapsed since the end of the previous trading day, and $\gamma_i > 0$ $(i = 1, ..., 8)$.[38] The $\gamma_i$ $(i = 2, ..., 7)$ are the values of the deterministic component $(i - 1.5)$ hours into each trading day. Note that we allow $\gamma_1 \neq \gamma_8$.[39]

MLEs and diagnostic statistics for the (unrestricted) g-HawkesE(2) model and the g-HawkesE(2) model with the restrictions $\pi_j = 0$ $(j = 1, 2)$ imposed are presented for both stocks in Table 2. The latter model imposes a lack of dependence between trading days and is thus labelled as the 'No dependence' model. LR tests of the restrictions $\pi_j = 1, \rho_j = \beta_j$ $\forall j$ (which yield the HawkesE(2) model given by (18)) are also reported and are compared to the $\chi^2(4)$ distribution. For each parameter of the selected models, non-symmetric 95% confidence intervals are reported in square parentheses.[40] In order to aid an intuitive understanding of the model, Figure 1 shows

---

[38] $\lfloor x \rfloor$, the integer part of $x$, denotes the greatest integer that is less than or equal to $x$.

[39] It is convenient to parametrise $\mu(t)$ by the $\gamma_i$ rather than in terms of the slopes of the linear pieces since this makes it straightforward to impose the constraint $\mu(t) > 0$ when performing the numerical optimisation.

[40] The confidence intervals (CIs) were calculated as follows. Letting $\theta$ denote the parameters of the selected

|  | **GENZ** |  | **DIS** |  |
| Model | g-HawkesE(2) | No dependence | g-HawkesE(2) | No dependence |
| --- | --- | --- | --- | --- |
| $\alpha_1$ | 98.043 [95,102] | 99.125 | 3.7671 | 3.7773 [3.2,4.5] |
| $\alpha_2$ | 1.7217 [1.4,2.1] | 1.3053 | 0.2570 | 0.2622 [0.14,0.49] |
| $\beta_1$ | 190.61 [182,199] | 188.24 | 9.5658 | 9.6078 [7.2,12.8] |
| $\beta_2$ | 5.0319 [4.2,6.1] | 3. 8743 | 0.4544 | 0.4611 [0.21,0.99] |
| $\pi_1$ | 1.4834 [0.84,2.6] | - | 0.0351 | - |
| $\rho_1$ | 152.64 [87,269] | - | 0.2504 | - |
| $\pi_2$ | 1.9066 [1.2,2.9] | - | 0.0340 | - |
| $\rho_2$ | 20.076 [12,33] | - | 0.0858 | - |
| $\gamma_1$ | 122.27 [95,157] | 218.17 | 170.40 | 177.75 [165,192] |
| $\gamma_2$ | 52.417 [42,65] | 45.058 | 96.660 | 103.12 [90,118] |
| $\gamma_3$ | 33.443 [26,44] | 31.369 | 42.913 | 47.863 [35,66] |
| $\gamma_4$ | 26.527 [20,34] | 23.450 | 13.468 | 17.557 [8.4,37] |
| $\gamma_5$ | 25.201 [20,32] | 23.510 | 0.4980 | 4.0404 [0.19,85] |
| $\gamma_6$ | 29.438 [24,36] | 28.146 | 14.331 | 17.383 [7.3,41] |
| $\gamma_7$ | 35.560 [29,43] | 34.624 | 24.289 | 26.836 [14,52] |
| $\gamma_8$ | 90.324 [79,103] | 91.106 | 81.087 | 83.084 [64,107] |
|  |  |  |  |  |
| log-likelihood | 335,859 | 335,767 | 194,659 | 194,659 |
| $LR$ ($\pi_j = 1, \rho_j = \beta_j$) | **60.180** (0.000) | - | **119.21** (0.000) | - |
|  |  |  |  |  |
| *Diagnostics* |  |  |  |  |
| Mean | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Variance | 0.9689 | 0.9694 | 1.0064 | 1.0058 |
| $BL$ | 18.244 (0.250) | **41.746** (0.000) | **31.182** (0.008) | **31.249** (0.008) |
| $BL$ (squares) | 22.800 (0.088) | **49.292** (0.000) | 6.1732 (0.977) | 6.1618 (0.977) |
| $ED$ | **-2.9091** (0.004) | **-2.8618** (0.004) | 0.4832 (0.966) | 0.4406 (0.882) |

Table 2: Generalised HawkesE(2) Models fitted to the times of trades for Genzyme (GENZ) and Walt Disney Company (DIS). The 'No dependence' model is the g-HawkesE(2) model with the restrictions $\pi_j = 0$ ($j = 1, 2$) imposed. 95% confidence intervals for the parameters of the selected models are shown in square parentheses. LR denotes a likelihood ratio test of the indicated restrictions. The mean and variance of the series on which the diagnostic tests are based are shown; BL is the Box-Ljung statistic for zero autocorrelation calculated using the first 15 lags; ED is the Engle and Russell (1998) test for excess dispersion; the diagnostic tests are described in detail in Section 3.2.

the various components of the estimated intensity for the GENZ stock on one randomly selected trading day (namely, 16 March 2000). The main features of the results are as follows.

In the case of GENZ our preferred model is the g-HawkesE(2) model, whereas we tentatively accept the restrictions $\pi_j = 0$ ($j = 1, 2$) in the case of DIS. The results of the diagnostic tests and issues related to testing these restrictions are discussed below. Consideration of the MLEs of ($\alpha_1, \beta_1, \alpha_2, \beta_2$) for the selected models reveals that, in both cases, the initial response of the

model shown in Table 2, the log-likelihood was first parametrised by $\phi = \log \theta$ and a 95% CI for each element of $\phi$ was formed in the usual way using the inverse of the negative Hessian matrix. The end points of these CIs were then exponentiated to obtain 95% CIs for $\theta$, thus ensuring that the final CI contains only positive values.
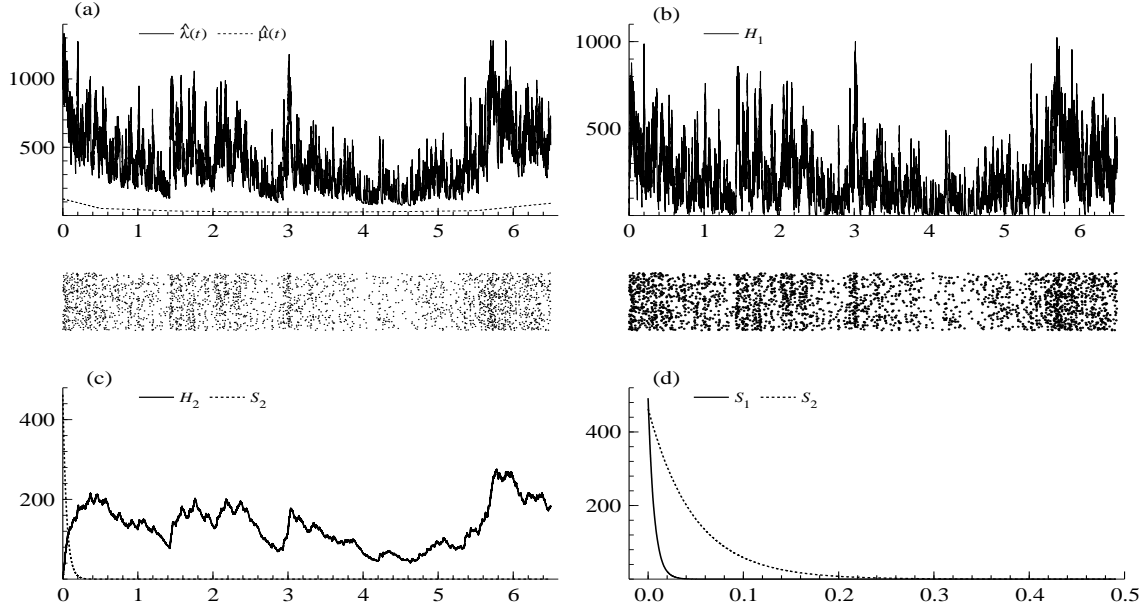
Figure 1: Components of the estimated trade intensity of Genzyme General during 16 March 2000. (a) the estimated total intensity, $\hat{\lambda}(t)$, and estimated deterministic component, $\hat{\mu}(t)$; (b) $H_1$, the estimated Hawkes-type component of $\tilde{\lambda}_1(t)$; (c) $S_2$ and $H_2$, the estimated Hawkes-type and spillover components of $\tilde{\lambda}_2(t)$; (d) the $j = 1$ and $j = 2$ spillover components plotted during the first 0.5 trading hours. Also shown in both columns are the (vertically jittered) occurrence times of the trades. In all cases, the horizontal axis is time measured in hours.

intensity to a trade is mainly the result of the first exponential response function ($j = 1$) but that this component decays rapidly whilst the second response function ($j = 2$) is initially smaller but also decays much more slowly. The second function could be referred to as the 'long lived' response function. The deterministic components exhibit the 'U-shaped' pattern familiar from earlier studies (see, for example, Engle and Russell (1998) and Engle (2000)), with high intensity around the open and close, and the lowest intensity occurring around 13:00.[41]

These features can be seen in panels (a)-(c) of Figure 1 for GENZ, which show: (a) the total estimated intensity, $\hat{\lambda}(t)$, together with the estimated deterministic component, $\hat{\mu}(t)$; (b) $H_1$, the estimated Hawkes-type component of $\tilde{\lambda}_1(t)$ (see (17)); and (c) $H_2$ and $S_2$, the estimated spillover and Hawkes-type components of $\tilde{\lambda}_2(t)$. Also shown in both halves of the figure are the occurrence times of the trades, with the points randomly vertically 'jittered' to aid clarity. A

---

[41]We also estimated a g-HawkesE(2) model with a deterministic component, $\mu(t)$, that incorporates day of the week effects. Specifically, we set $\mu(t) = \mu_\gamma(t) \cdot \gamma_{day}$, where $\mu_\gamma(t)$ is again given by (23) and, for example, $\gamma_{day} = \gamma_{Mon}$ if $t$ occurs on a Monday. $\gamma_{Fri}$ was normalised to one. The LR tests for the reduction to the (unrestricted) g-HawkesE(2) models shown in Table 2 accept for both stocks at the 5% level and there was little difference in the diagnostics obtained for either stock.

striking feature is that $H_2$ is very much smoother than $H_1$. We also see that the deterministic component contributes a relatively small fraction of the total intensity.[42] We have already discussed the difficulties associated with testing 'boundary hypotheses' such as $H_0 : \alpha_2 = 0$. Nevertheless, it is worth noting that imposing the restriction $\alpha_2 = 0$ and estimating the resultant g-HawkesE(1) model results in a reduction in the log-likelihood of 466 and 30 for GENZ and DIS respectively. The g-HawkesE(3) model was also estimated for both stocks but the sum (over $j$) of the estimated response functions was exactly the same as for the g-HawkesE(2) model.[43]

Of particular interest here are the estimates of the parameters $(\pi_1, \rho_1, \pi_2, \rho_2)$ that govern the dependence between trading days and the effect of imposing the restrictions $\pi_j = 0$ $(j = 1, 2)$. Again, in the absence of the necessary distribution theory, we cannot formally test the boundary hypothesis $H_0 : \pi_j = 0$ $(j = 1, 2)$ and so some of the comments that follow are necessarily tentative. In the case of GENZ, imposing these restrictions results in a reduction of 92.4 in the log-likelihood and both Box-Ljung statistics having $p$-values of 0.000. An *a priori* modelling assumption of no dependence between trading days would be undesirable in this case. Panel (d) of Figure 1 shows the estimated $j = 1$ and $j = 2$ spillover effects during the first 0.5 trading hours of the selected day for GENZ. Whilst the decay is very rapid for $j = 1$, it takes about 15 minutes for the $j = 2$ spillover effect to decay to a value close to zero. Examining the results for DIS reveals that imposing the restrictions $\pi_j = 0$ $(j = 1, 2)$ causes the log-likelihood to decrease by only 0.7, has a negligible effect on the diagnostic statistics, and does not result in large differences between the restricted and unrestricted MLEs. Also, the adoption of the minimum AIC procedure (Akaike (1977)) would result in the selection of the 'No dependence' model (with AIC of -595,675 compared to -595,668) as a result of the gain in parsimony. This model is thus tentatively accepted in the case of DIS. It should be stressed that an empirical finding of an absence of spillover effects in certain settings does not render models that incorporate these effects less valuable. In principle, such models make it possible to *test* the hypothesis of a lack of dependence and the failure to reject such a hypothesis is interesting in and of itself.

---

[42]In order to provide a rough guide to the importance of including the deterministic component in the model specification (whilst ensuring the positivity of the intensity) we estimated the g-HawkesE(2) model using the GENZ data, with the restrictions $\gamma_1 = \gamma_2 = ... = \gamma_8 = 0.001$ imposed. The log likelihood decreased by 210 and the diagnostic statistics were $(39.7, 68.4, -1.97)$ for $(BL, BL \text{ (squares)}, ED)$ respectively.

[43]This finding appeared to be insensitive to the starting values used for numerical optimisation.

We note that the LR test of the restrictions implied by the HawkesE(2) model has a $p$-value of 0.000 for GENZ and that the restrictions result in an additional diagnostic test rejecting at the 5% level (the Box-Ljung test of the squares now has a $p$-value of 0.037). The particularly restrictive form of the spillover effects implied by the HawkesE(2) model is thus rejected for GENZ, again emphasising the benefits of estimating these effects from the data.

The diagnostic tests for the selected models are very well behaved indeed, especially given the size of the datasets used. The smallest $p$-value was that of the excess dispersion test for GENZ (0.004) and that of the Box-Ljung test for DIS (0.008). The selected models also represent a huge improvement over the fit of the non-homogeneous Poisson model to these datasets. This model is given by $\lambda_\theta(t) = \mu_\gamma(t)$, where $\mu_\gamma(t)$ is as in (23). The analogous diagnostic test statistics $(BL, BL \text{ (squares)}, ED)$ for the non-homogeneous Poisson models are $(10571, 2065.0, 117.90)$ and $(539.70, 3.9096, 10.358)$ for GENZ and DIS respectively. The g-HawkesE(2) model thus describes well the self-exciting aspect of the dynamics of the processes, in addition to the deterministic, intradaily seasonality.

Overall, the results demonstrate the advantages of using a sufficiently general initial model that allows the nature of the dependence between trading days to be estimated from the data. As is apparent here, no single *a priori* modelling assumption is valid for all stocks.[44] We have considered a particular specification in which the $j$th spillover effect depends on $\tilde{\lambda}_j(\tau_{d-1})$, the level of the $j$th component of the intensity at the end of the previous trading day. In an alternative specification that may prove useful the spillover effect depends on the entire path of the intensity during the previous day via the term $\int_{\tau_{d-2}}^{\tau_{d-1}} W(\tau_{d-1} - s)\tilde{\lambda}_j(s)ds$, where $W(.) \geq 0$ is a non-negative 'weighting' function. The model structure introduced here in which the non-deterministic components of the intensity on trading day $d$ are specified recursively in terms of functionals of the intensity paths on previous days and the contributions of the events occurring on day $d$ is very general. This structure provides a useful framework for approaching the issue of dependence between trading days in PP models of financial markets.

---

[44]The difference between the estimated spillover effects for GENZ and DIS may well be due to the different opening procedures on NASDAQ and at the NYSE.

# 5 Bivariate intensity-based models

Having discussed the basic 'building block' of our (univariate and bivariate) generalised Hawkes models – the $\tilde{\lambda}_j(t)$ in equation (17) above – we are now in a position to turn again to the main focus of the paper: namely, the specification of multivariate models for market event data. The bivariate g-HawkesE(k) model and likelihood are discussed below, drawing to a considerable extent on our earlier treatment of the univariate model. An application to the timing of trades and mid-quote changes for an NYSE stock is then presented and the empirical results are related to the existing theoretical and empirical market microstructure literature.

## 5.1 Bivariate g-HawkesE(k) models

The univariate generalised Hawkes models can readily be extended to the multivariate case by including terms that capture the effect of $q$-type events ($q \neq m$) on the $m$th stochastic intensity. In the bivariate (BV) case, equation (16) is replaced by

$$\lambda_m(t) = \mu_m(t) + \sum_{j=1}^{k} \tilde{\lambda}_{mm}^{(j)}(t) + \sum_{j=1}^{k} \tilde{\lambda}_{mq}^{(j)}(t), \tag{24}$$

for $m = 1, 2$, where $\mu_m(t)$ is a positive deterministic function, $q = 2$ if $m = 1$ and $q = 1$ otherwise, and

$$\tilde{\lambda}_{mr}^{(j)}(t) = \pi_{mr}^{(j)} \tilde{\lambda}_{mr}^{(j)}(\tau_{d-1}) e^{-\rho_{mr}^{(j)}(t-\tau_{d-1})} + \int_{[\tau_{d-1},t)} \alpha_{mr}^{(j)} e^{-\beta_{mr}^{(j)}(t-u)} dN_r(u), \tag{25}$$

for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ...)$, $\tilde{\lambda}_{mr}^{(j)}(0) = 0$ and $mr \in \{1, 2\} \times \{1, 2\}$. The now familiar parameter restrictions $\alpha_{mr}^{(j)} \geq 0, \beta_{mr}^{(j)} > 0, \pi_{mr}^{(j)} \geq 0, \rho_{mr}^{(j)} > 0$ $(\forall mr)$ apply. The crucial extension is the inclusion of the terms $\tilde{\lambda}_{mq}^{(j)}(t)$ in (24), which allow the occurrence of $q$-type events to influence the intensity for $m$-type events. Whilst the notation necessarily becomes more cumbersome in the bivariate case, the essential building block of the model has not changed, as is evident from a comparison of equations (25) and (17). We thus retain a key feature of the univariate model – that is, the flexible specification of intensity spillover effects from one trading day to the next. The model given by (24) and (25) is henceforth referred to as the BV-g-HawkesE(k) model. Analogously to the univariate case, we note two special cases of this model. First, under the restrictions ($\pi_{mr}^{(j)} = 1$ and $\rho_{mr}^{(j)} = \beta_{mr}^{(j)}$ $\forall j, \forall mr$), the BV-g-HawkesE(k) model is equivalent to the

'mutually exciting' model given in Hawkes (1971).[45] Second, when $\pi_{mr}^{(j)} = 0 \ \forall j$ and $\forall mr$ there is no dependence between days since there are then no intensity spillover effects between days for either $\lambda_1(t)$ or $\lambda_2(t)$.

Taking logarithms of (6) yields the following log-likelihood for the BV-g-HawkesE(k) model:

$$l(\theta) = l_1(\theta_1) + l_2(\theta_2), \tag{26}$$

where $\theta = (\theta_1, \theta_2)$, $\theta_m \in \Theta_m$ is the parameter vector of the $m$th stochastic intensity, and

$$l_m(\theta_m) = \sum_{d=1}^{T/l} \left\{ \int_{A_d} (1 - \lambda_m(s; \theta_m)) ds + \int_{A_d} \log \lambda_m(s; \theta_m) dN_m(s) \right\}, \tag{27}$$

where $\lambda_m(s; \theta_m)$ is given by (24) and (25), $A_d = (\tau_{d-1}, \tau_d]$, and we have decomposed $l_m(\theta_m)$ into the contributions of the different trading days as before. Writing $\arg\max_{\theta_m} l_m(\theta_m)$ as $\hat{\theta}_m$, it follows from the fact that $\theta_1$ and $\theta_2$ are variation free (i.e. $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_1$) that $\arg\max_\theta l(\theta) = (\hat{\theta}_1, \hat{\theta}_2)$.[46] We therefore obtain the MLE of $\theta$ by separate numerical optimisation of (27) for $m = 1$ and $m = 2$ in order to reduce the dimensionality of the parameter space of the optimisation problem.[47] Further computational details related to estimation of the BV-g-HawkesE(2) model are given in the Appendix.

## 5.2 Timing of trades and quote changes on the NYSE

The BV-g-HawkesE(2) model is now used to analyse the timing of trades and mid-quote changes occurring on the NYSE for the stock General Motors Corporation (GM).[48] This application is of economic interest since the bivariate model allows the study of the two way interaction between the arrival processes for trades and price changes, thus providing a microstructure view of the relationship between trading activity and price volatility.

---

[45]In line with our terminology for the univariate case, we refer to the BV-g-HawkesE(k) model with these parameter restrictions as the BV-HawkesE(k) model.

[46]If the parameters of interest are, say, the parameters of $\lambda_1(t)$ alone (denoted by $\theta_1$), and we have some bivariate PP model satisfying that $\theta_1$ and $\theta_2$ are variation free (as is the case here), then the MLE of $\theta_1$ is obtained from the conditional model $\lambda_1(t)$ and the associated 'conditional likelihood' $l_1(\theta_1)$. Under these conditions, we might say that $\{T_i^{(2)}\}$ is weakly exogenous for $\theta_1$.

[47]The numerical optimisation was again performed using the BFGS algorithm with numerical derivatives.

[48]The 'specialist' or 'market maker' for each stock on the NYSE is required to report the best quotes (i.e. the highest bid and lowest offer) communicated to 'the crowd' by the specialist himself or a floor broker. The 'mid-quote' is defined here as the simple average of the reported bid and ask quotes. The quotes may consist of any of the following: the specialist's own trading interest, the trading interest of floor brokers in the crowd, or limit orders in the specialist's Display Book. The specialist is obliged to execute any order at a price that is at least as favourable as his published quote. Details of the institutional features of the NYSE, including reporting procedures for trades and quotes, may be found in Hasbrouck, Sofianos, and Sosebee (1993).

The period analysed is the 40 trading days from 5 July 2000 to 29 August 2000. The data transformation described in Section 4.1 is used to map the times of all trades and all changes to the mid-quote occurring during normal trading hours onto $[0, \infty)$. The times of the mid-quote changes are then thinned (i.e. a subset of the times is selected) to obtain the 'mid-quote events' that we model. We define a mid-quote event as occurring at the earliest time that the mid-quote changes by an amount greater than or equal to \$1/16 (in absolute value terms) compared to the mid-quote in force at the time of the previous mid-quote event.[49] The aim of our analysis here is to examine at a microstructure level of detail the two way interaction of trades and market quotes. It is therefore desirable not to set the threshold used to thin the quote data too high, in order to capture changes to the mid-quote over short time horizons. The threshold of \$1/16 used here is approximately equal to half of the average spread of \$0.117, and thus represents a very small movement in the mid-quote.[50] In order to avoid the model becoming overly complicated, we do not directly model the size of the change in the mid-quote between successive events. Given the way the mid-quote events are defined, the absolute size of this change can exceed \$1/16. However, analysis of this dataset showed that the majority (82.4%) of these changes do indeed equal \$1/16.[51]

Recall from Definition 2 that, since the pooled process $\{T_i\}$ is always simple, the $M$-variate PPs that we deal with here assign zero probability to the simultaneous occurrence of two events (of either the same or different types). In the case of the GM dataset analysed here, approximately 11 per cent of the mid-quote events have exactly the same timestamp as a trade. Since the simultaneous occurrence of a trade and a mid-quote event in the data will almost always be the result of lags between the (non-simultaneous) actual occurrence times of the events in

---

[49] The first mid-quote event is defined to occur at the time when the first pair of quotes were reported for the first trading day in the data set. Denote by $\{T_1^{(1)}, T_2^{(1)}, ...\}$ the (transformed) times of all changes to the mid-quote. The first mid-quote event time is $T_1^{(1)}$; the second mid-quote event time is $\min\{T_i^{(1)} : i > 1, |q(T_i^{(1)}) - q(T_1^{(1)})| \geq 0.0625\}$, where $q(T_i^{(1)})$ is the mid-quote reported by the specialist at time $T_i^{(1)}$. Subsequent mid-quote events are defined similarly. Engle and Russell (1998) use an analogous procedure to define their price events.

[50] The fit of the model when all changes to the mid-quote are included in the data is somewhat worse but suggests that perhaps some quite minor change to the model specification (such as the use of a more flexible functional form for the response function) could rectify the problem. We have not pursued this further in the current paper.

[51] The absolute changes are always some multiple of \$1/32, with the proportion of the changes equal to \$2/32, \$3/32, \$4/32 being given by 82.4%, 12.9%, and 3.0% respectively. An equivalent analysis of the absolute value of *all* changes to the mid-quote for this data set reveals that 85.9%,11.1%,1.9% are equal to \$1/32,\$2/32,\$3/32 respectively.

|  | Trade durations | Mid-quote event durations |
|---|---|---|
| No. of durations | 33,372 | 5,044 |
| Mean duration | 0.467 | 3.087 |
| Std deviation | 0.108 | 4.505 |
| Minimum | 1/60 | 1/60 |
| Maximum | 28.733 | 60.800 |
| $BL$ | **4733.1** (0.000) | **1192.1** (0.000) |

Table 3: Summary statistics of the durations between trades and mid-quote events for General Motors Corporation. BL is the Box-Ljung statistic for zero autocorrelation calculated using the first 15 lags; the durations are measured in minutes.

continuous time and the reported times, we adjust the times of those mid-quote events that coincide with a trade by a small, *i.i.d.* uniform amount. This procedure retains the original sequence of the other events and leaves the occurrence time of the mid-quote event unchanged to the nearest second. The adjustment is uniformly distributed because we do not want to impose prior assumptions about the original ordering of the two events. A discussion of the adjustment procedure used and two possible alternatives is given in the Appendix, together with the results of a sensitivity analysis comparing the three adjustment rules. These results support the robustness of the main conclusions of our empirical findings.

Summary statistics of the intertrade durations and durations between successive mid-quote events are given in Table 3 for the GM data. The average duration between mid-quote events is approximately 6.6 times the average intertrade duration. The results of fitting a restricted BV-g-HawkesE(2) model with the restrictions ($\pi_{mm}^{(1)} = 0, \pi_{mq}^{(1)} = 0, \pi_{mq}^{(2)} = 0; m = 1, 2; q = 2$ if $m = 1$ and $q = 1$ otherwise) are shown in Table 4.[52] The deterministic components, $\mu_m(t; \gamma_{m1}, ..., \gamma_{m8})$, have exactly the same form as in equation (23). The restrictions imposed imply that the non-deterministic components of the intensity exhibit zero spillover effects, except for the $\tilde{\lambda}_{mm}^{(2)}(t)$ components ($m = 1, 2$) which capture the $j = 2$ effect of $m$-type events on the $m$th intensity. The restricted model fits very well indeed, with only one of the diagnostic tests rejecting at the 5% level.[53] Figure 2 graphs the estimated components of the mid-quote event and trade

[52]The results of fitting the unrestricted BV-g-HawkesE(2) model are not shown here since the unrestricted model was found to have identical log-likelihood and diagnostic statistics. The 'restricted parameters' had MLEs very close to zero in the unrestricted model and the 'unrestricted parameters' were virtually identical for the two models.

[53]Two sets of diagnostics are reported in Table 4: one for the quote intensity and one for the trade intensity. The ($BL, BL$(squares), $ED$) tests based on the durations of the pooled process (formed *after* the changes of time) were ($23.670, 23.277, -3.7246$), again indicating a quite pleasing goodness of fit. The mean and variance of these durations were 0.98 and 0.90. The reader is referred to Section 3.2 for details of the diagnostics used.

| $\boldsymbol{\theta}_1$ | Quote intensity - $\lambda_1(t;\theta_1)$ | Trade intensity - $\lambda_2(t;\theta_2)$ | $\boldsymbol{\theta}_2$ |
|---|---|---|---|
| $\alpha_{11}^{(1)}$ | 1.7224 [1.2,2.5] | 10.388 [8.9,12] | $\alpha_{22}^{(1)}$ |
| $\alpha_{11}^{(2)}$ | 0.1469 [0.02,1.1] | 1.4838 [0.90,2.4] | $\alpha_{22}^{(2)}$ |
| $\beta_{11}^{(1)}$ | 8.6842 [5.3,14] | 42.716 [31,59] | $\beta_{22}^{(1)}$ |
| $\beta_{11}^{(2)}$ | 1.4072 [0.41,4.9] | 3.2561 [2.1,5.0] | $\beta_{22}^{(2)}$ |
| | | | |
| $\pi_{11}^{(2)}$ | 1.7270 [0.07,43] | 0.2103 [0.05,0.84] | $\pi_{22}^{(2)}$ |
| $\rho_{11}^{(2)}$ | 4.0172 [0.55,29] | 1.3812 [0.50,3.8] | $\rho_{22}^{(2)}$ |
| | | | |
| $\alpha_{12}^{(1)}$ | 41.009 [37,46] | 216.82 [180,261] | $\alpha_{21}^{(1)}$ |
| $\alpha_{12}^{(2)}$ | 4.4637 [2.7,7.3] | 38.830 [27,55] | $\alpha_{21}^{(2)}$ |
| $\beta_{12}^{(1)}$ | 810.43 [677,971] | 1550.8 [1155,2083] | $\beta_{21}^{(1)}$ |
| $\beta_{12}^{(2)}$ | 96.386 [64,146] | 101.42 [70,148] | $\beta_{21}^{(2)}$ |
| | | | |
| $\gamma_{11}$ | 21.740 [14,34] | 74.077 [53,104] | $\gamma_{21}$ |
| $\gamma_{12}$ | 2.6181 [0.76,9.0] | 45.906 [33,63] | $\gamma_{22}$ |
| $\gamma_{13}$ | 0.0000 | 15.580 [7.9,31] | $\gamma_{23}$ |
| $\gamma_{14}$ | 0.0000 | 21.365 [15,30] | $\gamma_{24}$ |
| $\gamma_{15}$ | 0.0000 | 17.928 [13,25] | $\gamma_{25}$ |
| $\gamma_{16}$ | 0.0000 | 26.408 [21,33] | $\gamma_{26}$ |
| $\gamma_{17}$ | 0.7743 [0.14,4.4] | 31.456 [25,39] | $\gamma_{27}$ |
| $\gamma_{18}$ | 0.4877 [0.01,37] | 69.602 [60,80] | $\gamma_{28}$ |
| | | | |
| $l_1(\theta_1)$ | 11,762 | 131,753 | $l_2(\theta_2)$ |
| | | | |
| Mean$^{(1)}$ | 0.9993 | 1.0000 | Mean$^{(2)}$ |
| Var$^{(1)}$ | 0.9741 | 1.0189 | Var$^{(2)}$ |
| $BL^{(1)}$ | 20.244 (0.163) | **46.487** (0.000) | $BL^{(2)}$ |
| $BL^{(1)}$ (squares) | 12.313 (0.655) | 4.7585 (0.994) | $BL^{(2)}$ (squares) |
| $ED^{(1)}$ | -0.6499 (0.516) | 1.2216 (0.222) | $ED^{(2)}$ |

Table 4: MLEs and diagnostics for a restricted BV-g-HawkesE(2) model of the timing of trades and mid-quote changes of General Motors Corporation. The parameters of the quote intensity and the trade intensity are listed in the first and last columns respectively. 95% confidence intervals are shown in square parentheses. The maximised log likelihood for the bivariate model is $11,762 + 131,753 = 143,515$. The reported diagnostics are analogous to those reported for the univariate models earlier and are described in Section 3.2; BL again denotes the Box-Ljung test and ED the excess dispersion test; a superscript (1) denotes a diagnostic based on the quote intensity and a (2) denotes one based on the trade intensity.

intensities (in panels (a),(c),(e) and (b),(d),(f) respectively) for a randomly selected trading day – day 21 of the dataset, i.e. 2 August 2000. The first panel of the $m$th column ($m = 1, 2$) shows the estimated total intensity, $\hat{\lambda}_m(t)$, and estimated deterministic component, $\hat{\mu}_m(t)$; the second shows the occurrence times of the $q$-type events ($q \neq m$) in order to highlight the impact of these events on $\hat{\lambda}_m(t)$; the third shows $\hat{\lambda}_{mq}(t)$ ($q \neq m$), which is the estimate of $\sum_{j=1}^{2} \tilde{\lambda}_{mq}^{(j)}(t)$ in (24); and the fourth shows $\hat{\lambda}_{mm}(t)$, the estimate of $\sum_{j=1}^{2} \tilde{\lambda}_{mm}^{(j)}(t)$, together with $S_{mm}$, the estimate of
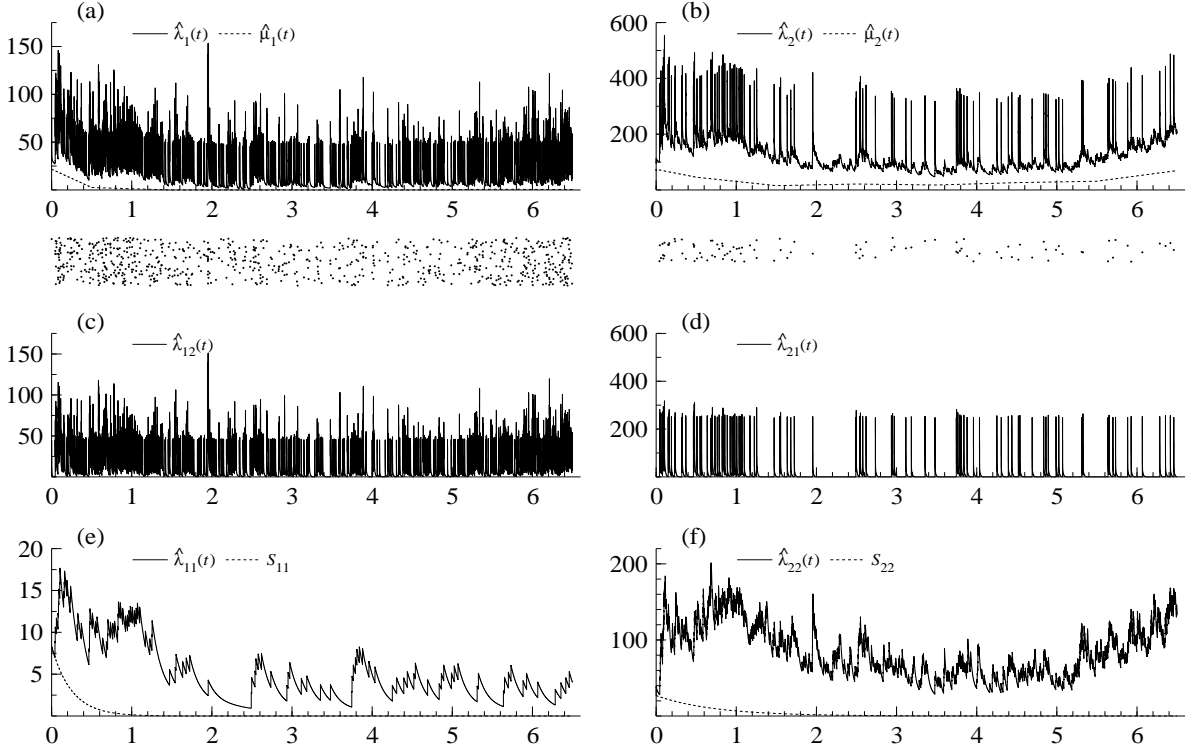
Figure 2: Components of the estimated mid-quote event and trade intensities for General Motors on 2 August 2000. The first column of panels is for the mid-quote event intensity and the second for the trade intensity. (a) the estimated total intensity, $\hat{\lambda}_1(t)$, and estimated deterministic component, $\hat{\mu}_1(t)$; (c) $\hat{\lambda}_{12}(t)$, the estimate of $\sum_{j=1}^2 \tilde{\lambda}_{12}^{(j)}(t)$; (e)$\hat{\lambda}_{11}(t)$, the estimate of $\sum_{j=1}^2 \tilde{\lambda}_{11}^{(j)}(t)$, and $S_{11}$, the estimate of the spillover effect $\pi_{11}^{(2)}\tilde{\lambda}_{11}^{(2)}(\tau_{20})\mathrm{e}^{-\rho_{11}^{(2)}(t-\tau_{20})}$. The other components shown are defined analogously. Also shown in the first (resp. second) column are the (vertically jittered) occurrence times of the trades (resp. mid-quote events). Note that panels (a) and (c), and (b) and (d) are drawn to the same scale. In all cases, the horizontal axis is time measured in hours.

the spillover effect $\pi_{mm}^{(2)}\tilde{\lambda}_{mm}^{(2)}(\tau_{20})\mathrm{e}^{-\rho_{mm}^{(2)}(t-\tau_{20})}$.

Consider first the MLEs reported in Table 4. Non-symmetric 95% confidence intervals were calculated as for the univariate case and are again shown in square parentheses.[54] The 'U-shape' of the deterministic component of the trade intensity, and the long lived $j = 2$ response functions which govern the evolution of the $\tilde{\lambda}_{mm}^{(2)}(t)$ terms in (24) ($m = 1, 2$) are familiar from the univariate results. The deterministic component of the quote intensity is close to zero after

---

[54]The CIs were obtained by exponentiating the CIs for the elements of $\phi_m = \log(\theta_m)$ ($m = 1, 2$). The latter were based on the inverse of the negative Hessian matrix, as in the univariate case. A pragmatic approach was taken in order to avoid the problems associated with $\gamma_{13}, \gamma_{14}, \gamma_{15}$ and $\gamma_{16}$ having values near to the boundary of the parameter space: the model was estimated with the additional restrictions $\gamma_{13} = \gamma_{14} = \gamma_{15} = \gamma_{16} = 0.0001$ imposed (in order to ensure positivity of the intensity process) and CIs for the other quote intensity parameters were obtained using these restricted estimates. Imposing the additional restrictions had virtually no effect on the log likelihood and MLEs for the quote intensity.

the first hour of the day. Of particular economic interest are the estimates of $\alpha_{mq}^{(1)}, \beta_{mq}^{(1)}, \alpha_{mq}^{(2)},$ and $\beta_{mq}^{(2)}$ $(m = 1, 2; q \neq m)$. We see that the occurrence of a trade results in an upward jump in the estimated mid-quote event intensity (equal to $\hat{\alpha}_{12}^{(1)} + \hat{\alpha}_{12}^{(2)}$) and that the effect then decays away with time. Similarly, the occurrence of a mid-quote event results in an increase in the estimated trade intensity. These effects bring about the large, short lived spikes that are evident in the estimated total intensities $\hat{\lambda}_m(t)$ $(m = 1, 2)$ in Figure 2. Notice that the magnitude of the spikes in $\hat{\lambda}_{12}(t)$ and $\hat{\lambda}_{21}(t)$ are large relative to the levels of $\hat{\lambda}_{11}(t)$ and $\hat{\lambda}_{22}(t)$ respectively. This is particularly pronounced in the case of the mid-quote intensity. The response functions of the 'cross effect' terms in (24), $\sum_{j=1}^{2} \tilde{\lambda}_{mq}^{(j)}(t)$, are very short lived, with the $j = 1$ component having a half life of 3.1 seconds in the case of $\hat{\lambda}_{12}(t)$ and 1.6 seconds in the case of $\hat{\lambda}_{21}(t)$. The picture that emerges from Figure 2 is one in which the cross effect terms which capture the effect of $q$-type events on the $m$th intensity $(q \neq m)$ exhibit large, short lived fluctuations and play an extremely important role in determining the dynamics of the process. We shall consider the economic interpretation of these effects below.

The hypothesis $H_0 : \alpha_{12}^{(j)} = 0$ $(j = 1, 2)$ corresponds to the case where the mid-quote event intensity does not depend on the history of trades. Again, since the parameter values under the null lie on the boundary of the maintained hypothesis and there are nuisance parameters that are unidentified under the null, the standard LR test cannot be applied. Nevertheless, we note that in this case imposing the restrictions $\alpha_{12}^{(j)} = 0$ $(j = 1, 2)$ on the $\lambda_1(t)$ intensity of the (unrestricted) BV-g-HawkesE(2) model yields a univariate g-HawkesE(2) model with a log-likelihood of 10,672 – a sizeable reduction of 1090 when compared to $l_1(\theta_1)$ in Table 4. All 3 diagnostic tests continue to accept at the 5% level. Similarly, imposing the restrictions $\alpha_{21}^{(j)} = 0$ $(j = 1, 2)$ on $\lambda_2(t)$ results in a reduction in the log-likelihood of 450 and in the excess dispersion test rejecting at the 1% level. It would seem that in this case, the bivariate model is a substantial improvement over the two univariate models which ignore the cross effects. Interestingly, this appears to be true not only of the mid-quote event intensity but also of the trade intensity, indicating that the effect of mid-quote events on the intensity for trades should not be ignored. A sensitivity analysis is presented in the Appendix comparing three different adjustment rules for the treatment of the simultaneous occurrence of trades and mid-quote events in the original
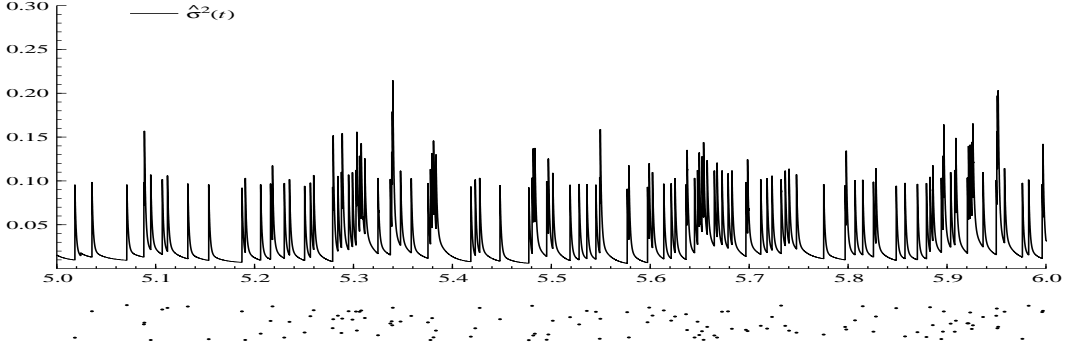
Figure 3: Estimated instantaneous volatility, $\hat{\sigma}^2(t)$, for General Motors on 2 August 2000 between 14:30 and 15:30 EST. $\hat{\sigma}^2(t)$ is based on equation (29). Also shown are the (vertically jittered) occurrence times of the trades. Volatility is measured per trading year (where one year is 252×6.5 trading hours). The horizontal axis is time measured in hours since the start of the trading day.

data. The analysis strongly suggests that our finding of a positive effect of mid-quote events on the trade intensity is not the result of the particular adjustment rule employed here.

An aim of one strand of empirical microstructure research is to investigate the relationship between the trade arrival process and volatility. In order to see how the mid-quote intensity of the BV-g-HawkesE(2) model can be used to obtain (an approximation to) the instantaneous volatility of the price process, consider the case where all changes to the mid-quote take values in $\{-c, +c\}$ and a mid-quote event is defined to occur whenever the mid-quote changes. Denote the time of the $i$th mid-quote event as usual by $T_i^{(1)}$ and the associated mid-quote change by $\Delta_i$. Then define the instantaneous conditional volatility by

$$\sigma^2(t) = \lim_{h \downarrow 0} \mathrm{E}\left[ \frac{1}{h} \left( \frac{P(t+h) - P(t)}{P(t)} \right)^2 | \mathcal{F}_t \right], \tag{28}$$

where $\mathcal{F}_t = \sigma(P(t)) \vee \mathcal{F}_t^N$, $N$ is the bivariate PP of trades and mid-quote events, and the right-continuous price process is given by $P(t) = P(0) + \sum_{i:T_i^{(1)} \leq t} \Delta_i$. Then, as Engle and Russell (1998) have noted, it is possible to express $\sigma^2(t)$ in terms of the mid-quote intensity. Specifically,

$$\sigma^2(t) = c^2 \lambda_1(t+)/P(t)^2, \tag{29}$$

where we have used Lemma 3.3(ii) of Aalen (1978) and have assumed that $\lambda_1(t)$ is the mid-quote event intensity with respect to $\mathcal{F}_t$ as well as $\mathcal{F}_t^N$ (i.e. conditioning additionally on the current price level does not alter the intensity).

35

An approximate estimate of the instantaneous volatility can be formed using the estimates for the restricted BV-g-HawkesE(2) model obtained above by making the following substitutions in (29): replace $\lambda_1(t+)$ by the (right-continuous version of the) estimated intensity $\hat{\lambda}_1(t)$, set $c = \$1/16$ and let $P(t) = P(0) + \sum_{i:T_i^{(1)} \leq t} \Delta_i$ (where $T_i^{(1)}$ is the time of the $i$th mid-quote event for the dataset and $\Delta_i$ is the actual change to the mid-quote since the last mid-quote event).[55] This estimate is graphed in Figure 3 for the trading hour between 14:30 and 15:30 EST on 2 August 2000 (the same day that was used in Figure 2) . Showing just one hour in this way allows the detail of the function and its relation to the timing of the trades to be clearly seen. A prominent feature is the association of periods of high volatility with periods of high trading activity (so called 'clusters' of trades). As expected, this feature was evident for all of the trading days that we graphed. Note also that the volatility and mid-quote intensity estimates (the latter not shown here) were difficult to distinguish as a result of the relatively small variability of the price compared to that of the intensity. The BV-g-HawkesE(2) model can thus be used to obtain a microstructure view of the interaction between financial market volatility and its determinants. It would be interesting to condition also on the volume of trades and the size and direction of the price changes between mid-quote events. The extension of our approach to models that are conditional on a filtration wider than the internal history of the PP is discussed in Section 6.3.2 below.

## 5.3  Market Microstructure

Much of the existing theoretical and empirical market microstructure literature concentrates on the impact of trades on prices rather than on the reciprocal effect of prices on the trade arrival process. By contrast, the modelling framework presented here allows the *two way* interaction of trades and mid-quote changes to be analysed. The empirical results of Section 5.2 are related to the theoretical and empirical microstructure literatures in turn below.

The theoretical literature is concerned with the role that the trading process plays in the formation of security prices, and in particular with how new information is incorporated into

---

[55]There are 2 sources of approximation error here. First, the mid-quote is assumed to change only at the times of the mid-quote events whereas in reality it changes more frequently. Second, (29) holds exactly when $\Delta_i^2 = c^2$ $\forall i$ but we observe $\Delta_i^2 > c^2$ for a minority of mid-quote events in the dataset. Nevertheless, the approximation is a useful one.

prices.[56] An important class of models is the sequential trade models of Glosten and Milgrom (1985), Easley and O'Hara (1987) and Easley and O'Hara (1992). In these models, the Bayesian specialist learns about the information held by the informed traders by observing the sequence of trade outcomes and sets his quotes equal to the expected terminal value of the asset conditional on the past sequence of trading outcomes and a trade at the quote. Thus, the dynamics of the posted quotes and of transaction prices result from this Bayesian updating. Crucially, trades convey information to the specialist and so impact the quoted prices. A central feature of the Easley and O'Hara (1992) model is that uninformed market participants, including the specialist, are uncertain whether an information event has occurred prior to the start of a given trading day. This results in periods of low volatility tending to occur in periods when there are few trades, since such a period increases the probability the specialist attaches to there having been no information event at the start of the trading day.

The finding in Section 5.2 above that the occurrence of a trade results in an increase in the mid-quote intensity is thus consistent with the central feature of the sequential trade models: namely, that the specialist updates his beliefs about the value of the stock in response to the trades that he observes. A change to the mid-quote is thus more likely immediately following a trade. We note that a trade would also trigger a mid-quote change when one of the sides to the trade was a limit order that constituted the market quote before the trade took place. Theoretical microstructure models have analysed in much less detail the effects of quoted prices on the trade arrival process and so it is more difficult to interpret our finding that the occurrence of a mid-quote change results in an increase in the trade intensity. A broad explanation is that some market participants closely monitor the quoted prices and rapidly submit market orders in order to take advantage of prices that are favourable to them, whilst others may set their quotes in order to attract such market orders. For example, when the mid-quote change is the result of inventory control by the specialist (see Ch. 2, O'Hara (1995)), the altered quote will tend to be 'hit' soon afterwards by an attracted market order on the opposite side of the market. Asymmetric information considerations might predict a longer run negative effect of quote changes on trading intensity, reflecting the incorporation of private information into the

---

[56]O'Hara (1995) and Hasbrouck (1996) give book and review length treatments respectively of this literature.

stock price over time and hence reduced incentives for the informed to trade. Such an effect is not possible in our model since the constraints $\alpha_{21}^{(j)} \geq 0 \; \forall j$ are imposed. As is discussed briefly in Section 6.3.2 it is possible to allow for such negative effects in our modelling framework and such an extension would be of some interest given our findings here.

In our model, a cluster of trades with short intertrade durations results in a large increase in the mid-quote event intensity and thus a large increase in volatility (see Figure 3). This is consistent with the central prediction of the Easley and O'Hara (1992) model noted above – namely, that periods of low volatility tend to occur in periods when there is little trade. A number of other empirical studies have reported similar findings.[57]

With the exception of Engle and Lunde (2003), none of these papers model the dependence of the intertrade durations on prices. Both Russell and Engle (1998) and Engle (2000) assume that intertrade durations are not Granger caused by prices. In contrast, the bivariate modelling approach adopted here is ideally suited to the study of the two way interaction between trades and prices. Dufour and Engle (2000) provide preliminary evidence that short intertrade durations tend to follow durations with large absolute returns (and large trades). This is in line with our finding that the occurrence of a mid-quote change results in a large increase in the trade intensity. In contrast, Grammig and Wellner (2002) finds that higher lagged volatility significantly reduces trade intensity.[58] The further investigation of the effect of prices on the trade arrival process is an interesting challenge for both theoretical and empirical microstructure research.

---

[57]Engle and Lunde (2003) found that short intertrade durations predict short (observed forward) mid-quote event durations; Engle and Russell (1998) found that expected price durations were shorter following price durations with a higher number of trades per second; Engle (2000) reports that the conditional volatility per unit time over an intertrade duration was higher when both the expected and actual duration were shorter; and Russell and Engle (1998) note that the expected squared price change over an intertrade duration was a decreasing function of the expected length of that duration.

[58]Note, however, that this study analysed data for a stock on an electronic limit order book system in the five weeks following a large initial public offering, a situation in which asymmetric information effects would be expected to be particularly prevalent.

# 6   Comparison with previous research and possible extensions

## 6.1   Model specification via durations

Engle and Lunde (2003) (hereafter EL) specify models for a bivariate sequence of durations $\{S_i^{(1)}, S_i^{(2)}\}_{i \in \{1,2,...\}}$, where $S_i^{(m)} = T_i^{(m)} - T_{i-1}^{(m)}$ is the $i$th successive duration between the $m$-type events. The events are defined so that $S_i^{(1)}$ and $S_i^{(2)}$ refer to time intervals with a common origin in real time. In their setting, $S_i^{(1)}$ is the $i$th intertrade duration and $S_i^{(2)}$ is the time from the start of the $i$th intertrade duration until the next trade or mid-quote change (whichever occurs first). The trade arrival process is thus the 'driving process' and the arrival processes for trades and mid-quote changes are not treated symmetrically. This approach avoids the difficulty that arises in general when constructing bivariate models via conventional time series modelling of $\{S_i^{(1)}, S_i^{(2)}\}$: namely, that the origins of $S_i^{(1)}$ and $S_i^{(2)}$ may well be far apart in real time.

The EL model is a rather specialised model, designed with a particular economic question in mind – that is, the speed with which the information content of trades is incorporated into prices. This sharp focus does, however, limit the scope and wider applicability of the model. Indeed, the EL model does not (and does not claim to) constitute a full bivariate PP model. In particular, it does not imply an intensity in continuous time for mid-quote change events. Other restrictive features are that the occurrence of a mid-quote change during an intertrade duration cannot influence the trade intensity during that duration, and that there is an implicit loss of information when multiple mid-quote changes occur without an intervening trade. Also, the selection of a single 'driving process' will often be difficult to justify on economic grounds when moving to higher dimensional multivariate PP models.

The above discussion prompts the question of how a full $M$-variate PP model could be specified via the conditional distribution of durations. We suggest specifying

$$\mathrm{P}[S_{i+1} \leq x, Z_{i+1} = m | \mathcal{F}_{T_i}^N] = \int_0^x f_{i+1}(s, m) \, ds, \tag{30}$$

for $m = 1, ..., M$, where $S_{i+1} = T_{i+1} - T_i$ is the $(i+1)$th duration of the pooled process $\{T_i\}$, $\mathcal{F}_{T_i}^N = \sigma(T_1, Z_1, T_2, Z_2, .., T_i, Z_i)$ and $\sum_{m=1}^M \int_0^\infty f_{i+1}(s, m) ds = 1$. That is, we suggest specifying the conditional density of the time to the next event of the pooled process given the multivariate

filtration at the start of the duration and given that the next event has type $m$.[59] Such an approach may prove useful in future work.[60]

## 6.2 Russell (1999)

Russell (1999) also adopts an intensity based approach. The Autoregressive Conditional Intensity (ACI) family of models and the suggestions of Russell (1999) concerning the construction of diagnostic tests are considered in turn below.

### 6.2.1 The Autorgeressive Conditional Intensity Model

We define here the bivariate ACI(1,1) model (with time varying intensity between events) that is estimated in the empirical section of Russell (1999).[61] In our terminology, this is an $M$-variate PP with $M = 2$. The $\mathcal{F}_t^N$-intensity of the process is given by

$$\lambda_m(t) = \mu_m(t) \exp[\phi_m(t)][U_m(t)]^{\varphi_{mm}} [U_n(t)]^{\varphi_{mn}}, \tag{31}$$

for $m = 1, 2$, where $\mu_m(t)$ is a positive deterministic function, $n = 2$ if $m = 1$ and $n = 1$ otherwise, $\omega_m, \varphi_{mm}, \varphi_{mn}$ are parameters, $U_q(t) = t - T_{\check{N}_q(t)}$ is the so called backward recurrence time for $q$-type events ($q = 1, 2$), and $\check{N}_q(t)$ is the process with *left continuous* sample paths that counts the $q$-type events. The central feature of the model is the $\phi_m(t)$ processes, whose sample paths are constant over the durations of the 'pooled process,' $\{T_i\}$. That is,

$$\phi_m(t) = \phi_{m,i} \quad \text{for } T_i < t \leq T_{i+1}. \tag{32}$$

An autoregressive specification is adopted for $\phi_i = (\phi_{1,i}, \phi_{2,i})'$, namely

$$\phi_i = \begin{cases} \eta_1 \varepsilon_i + B\phi_{i-1} & \text{if } Z_i = 1 \\ \eta_2 \varepsilon_i + B\phi_{i-1} & \text{if } Z_i = 2 \end{cases}, \tag{33}$$

where $\eta_1, \eta_2$ and $B$ are parameters of the appropriate dimension.[62] $\varepsilon_i$ is the scalar r.v. associated with the occurrence of the $i$th event of the pooled process. Let $T_i^*$ denote the time of the most

---

[59]Note that specifying $f_{i+1}(s, m)$ for $m = 1, ..., M$ and for all $i$ determines $(\lambda_m(t))_{m=1}^M$ uniquely.

[60]Indeed, we recently became aware of Kamionka (2001), which essentially adopts this approach to specify a bivariate PP model.

[61]The original notation of Russell (1999) has been adapted somewhat here, partly in order to highlight certain aspects of the model structure and partly to ensure consistent notation throughout this paper. In the notation used here the model might be better described as an ACI(1,0) model (using ARMA(p,q) type nomenclature).

[62]Notice that we have been careful to define the $\lambda_m(t)$ so that they have left continuous sample paths and are thus predictable. The use of right continuous $\lambda_m(t)$ in the expression for the likelihood in (6) is incorrect and introduces an error in the evaluation of the log-likelihood.

recent event that occurred prior to the $i$th event and is of the same type as that event. Formally, let $T_i^* = \sup\{T_j : j < i \text{ and } Z_j = Z_i\}$ provided that this set is not empty and let $T_i^* = 0$ otherwise. Then, the r.v. $\varepsilon_i(\omega)$ is given by

$$\varepsilon_i(\omega) = 1 - \tilde{\varepsilon}_i(\omega) = 1 - \int_{T_i^*(\omega)}^{T_i(\omega)} \lambda_{Z_i(\omega)}(\omega, s) ds. \tag{34}$$

In the univariate case, $T_i^* = T_{i-1}$ and $\phi_i = \eta \varepsilon_i + b \phi_{i-1}$, where $\{\varepsilon_i\}$ is an $i.i.d.$ sequence of mean zero r.v.'s (a fact that follows by setting $M = 1$ in Corollary 3.2 since $\lambda$ is the $\mathcal{F}_t^N$-intensity). $\{\phi_i\}$ is thus an AR(1) process. This way of specifying the model allows the powerful methods of time series analysis to be brought into play. However, we have been unable to prove that the $\varepsilon_i$ in (34) are independent (or even uncorrelated) in the bivariate case. It does *not* follow from Theorem 3.1 and Corollary 3.2 that $\{\varepsilon_i\}$ is an $i.i.d.$ sequence of mean zero r.v.'s in this case. Recalling the definition of the $\{e_i^{(m)}\}$ ($m = 1, 2$) series in Corollary 3.2, our notation here highlights that although a particular realisation, $\{\varepsilon_i(\omega)\}$, consists of some arrangement of the elements of $\{e_i^{(1)}(\omega)\}$ and $\{e_i^{(2)}(\omega)\}$, that arrangement is random (i.e. it varies with $\omega$). It may well be that further work will resolve this problem.[63] An alternative model is given by leaving the specification unchanged except that $\varepsilon_i$ is now given by

$$\varepsilon_i = 1 - \tilde{\varepsilon}_i = 1 - \int_{T_{i-1}}^{T_i} \sum_{m=1}^{2} \lambda_m(s) ds. \tag{35}$$

Since the $\mathcal{F}_t^N$-intensity of $\sum_m N_m(t)$ is $\sum_m \lambda_m(t)$, it now follows easily from arguments identical to those in Theorem 3.1 and Corollary 3.2 that $\{\tilde{\varepsilon}_i\}$ is $i.i.d.$ $Exp(1)$ in this case.

To sum up, the ACI-type model structure offers useful alternative model specifications to the generalised Hawkes models developed here. Given the complexity of financial market event data, the availability of different families of econometric models is to be welcomed. An advantage of the ACI framework is the rich variety of time series specifications that could be adopted for $\phi_i$. Possible disadvantages are the technical difficulties sometimes involved in establishing the properties of the $\tilde{\varepsilon}_i$ and the need to use numerical integration (in all but the simplest cases) to calculate the log-likelihood and diagnostic tests based on the integrated intensity. Ultimately,

---

[63]We would be grateful for any suggestions on proofs related to these points. In private correspondence, Jeffrey Russell has kindly informed me of the results of some simulation experiments which seem to support the conjecture that $\{\tilde{\varepsilon}_i\}$ is $i.i.d.$ $Exp(1)$.

both families of models will be judged by their ability to provide well specified empirical models whose parameters may be readily interpreted in economic terms.

### 6.2.2 Diagnostic testing revisited

We have commented already that Russell (1999) suggests the use of the $\{e_i^{(m)}(\theta_0)\}_i$ sequence defined by (15) as the basis for diagnostic testing and indicates that the results of Yashin and Arjas (1988) concerning the 'exponential formula' might be used to establish its *i.i.d.* $Exp(1)$ property, provided that a certain absolute continuity condition is satisfied. The difficulty in this context in applying the formula is that the $\sigma$-field upon which conditioning takes place is not fixed. Consider the process $F_i^{(m)}(x) = \Pr[S_i^{(m)} \leq x | \mathcal{G}_x]$, where $\mathcal{G}_x$ is the observed history on which the assessment of the hazard is based and $S_i^{(m)} = T_i^{(m)} - T_{i-1}^{(m)}$ as above.[64] Since non $m$-type events can occur within this duration the conditioning is 'dynamic' (i.e. $\mathcal{G}_x$ depends on $x$). In order to apply the exponential formula in the Corollary of Yashin and Arjas (1988) it is necessary that $F_i^{(m)}$ is a continuous function of $x$ (with probability one). This is an undesirably restrictive condition in our context since it does not allow $F_i^{(m)}$ to exhibit jumps in response to the occurrence of $q$-type events ($q \neq m$). Such jumps would occur when the influence of these events on $S_i^{(m)}$ is 'strong,' which is a likely scenario here.

The random change of time approach adopted in Section 3.1 thus offers two particular advantages – as we have shown, it makes possible a rigorous proof that the sequence is *i.i.d.* $Exp(1)$ and that proof avoids reliance on restrictive conditions that are unlikely to be satisfied by realistic models of multivariate market event data.

### 6.3 Possible Extensions

We outline here a number of directions in which our own work may be extended. For the sake of continuity with the discussion of diagnostic testing above, we first discuss the issue of how to take account of the impact of parameter uncertainty on our diagnostic testing.

---

[64]Justifying the application of the Theorem and Corollary of Yashin and Arjas (1988) to yield an expression for $\Pr[S_i^{(m)} \leq x | \mathcal{G}_x]$ in terms of the integral of the $\mathcal{F}_t^N$-intensity for $m$-type events is not trivial and Russell (1999) does not provide details of the derivation. If the conditioning were on a fixed $\sigma$-field, the information observed since the $(i-1)$th $m$-type event would be excluded and so an intensity other than the $\mathcal{F}_t^N$-intensity $\lambda_m(t)$ would then be required as the integrand. Our purpose here is not to explore these issues in any depth but rather to point out the considerable drawback of the absolute continuity restriction.

### 6.3.1 Parameter uncertainty and diagnostic testing

Two alternative feasible procedures that take account of the parameter uncertainty correctly are presented below, the first of which is applicable only to univariate PP models.

Thompson (2001) considers tests of correct model specification for a sequence of scalar r.v.'s $\{S_i\}$ based on the fact that applying the 'integral transforms' given by the true conditional distribution functions yields an *i.i.d.* uniform series. Asymptotically exact critical values and upper bounds for the asymptotic critical values are provided that take into account the fact that the parameter has been estimated. In order to see how these results apply in a PP setting, let $S_i = T_i - T_{i-1}$ be the $i$th duration of a univariate PP with $\mathcal{F}_t^N$-intensity $\lambda(t)$. Then it is possible to show, provided that $\lambda(t)$ can be expressed in a certain form, that[65]

$$F_i(x,\theta) = \Pr[S_i \leq x|S_{i-1},...,S_0;\theta] = 1 - \exp\left(-\int_{T_{i-1}}^{T_{i-1}+x} \lambda(s;\theta)ds\right). \qquad (36)$$

Clearly, $\{F_i(S_i,\hat{\theta})\} = \{1 - \exp(-e_{i-1}(\hat{\theta}))\}$ (where $e_{i-1}(\hat{\theta})$ is given by (15) and $M = 1$).[66] Thus, provided that the PP model in question satisfies the relevant assumptions of Thompson (2001), the critical values provided there will be valid for the tests of uniformity and independence based on $\{1 - \exp(-e_{i-1}(\hat{\theta}))\}$. However, this approach is generally not feasible in the case of a multivariate PP model. For example in the bivariate case, letting $S_i^{(m)} = T_i^{(m)} - T_{i-1}^{(m)}$ as usual, the series of durations $\{S_i^{(1)}, S_i^{(2)}\}_{i\in\{1,2,...\}}$ can be readily constructed. The problem is that the calculation of objects such as $\Pr[S_i^{(m)} \leq x|S_{i-1}^{(2)}, S_{i-1}^{(1)}..., S_1^{(2)}, S_1^{(1)};\theta]$ and $\Pr[S_i^{(m)} \leq x|S_i^{(q)}, S_{i-1}^{(2)}, S_{i-1}^{(1)}..., S_1^{(2)}, S_1^{(1)};\theta]$ $(q \neq m)$ is usually infeasible because the distribution functions are not known in closed form. In the context of diagnostic testing for multivariate models, as in that of model specification, the switch in focus from durations to intensities is a very powerful one.

An alternative procedure (also noted by Thompson (2001)) that applies to both the univariate and multivariate cases is to minimise the diagnostic test statistic with respect to the parameters. Let $T^{(m)}(\theta)$ denote the test statistic evaluated at $\theta$ (obtained by calculating the

---

[65]The condition that $\lambda(t)$ must satisfy and the main part of the proof may be found in E5, p.63 of Brémaud (1981). (36) is an example of the 'exponential formula' referred to in Section 6.2.2 above.

[66]Note that it is immediate from Corollary 3.2 that $\{F_i(S_i,\theta_0)\} = \{1 - \exp(-e_{i-1}(\theta_0))\}$ is *i.i.d.* uniform on $(0,1)$.

statistic using the series $\{e_i^{(m)}(\theta)\}$). The null of correct specification is rejected if and only if

$$T^{(m)} = \min_{\theta \in \Theta} T^{(m)}(\theta), \tag{37}$$

exceeds the $(1 - \alpha)$ quantile of the null distribution of $T^{(m)}(\theta_0)$ (where $\theta_0$ is the true parameter value). Since $T^{(m)} \leq T^{(m)}(\theta_0)$ everywhere, the probability of falsely rejecting the null is bounded above by $\alpha$. The validity of the bound does not depend on the PP model satisfying any regularity conditions other than the one in (12) and so the procedure has widespread applicability. Furthermore, the bound is valid when $\theta_0$ is on the boundary of the parameter space, a case which is of some importance for generalised Hawkes models. We have found that the calculation of such test statistics is computationally feasible for datasets of the size used here and can yield tests with sensible properties.[67]

### 6.3.2 Conditioning on a wider filtration

An important strength of the modelling framework developed in this paper is the ease with which the approach can be extended to model the evolution of the multivariate PP conditional on information that is additional to the internal history of the process, $\mathcal{F}_t^N$. In this case the econometrician specifies the $\mathcal{F}_t$-intensity of $N_m(t)$ for $m = 1, ..., M$, where $\mathcal{F}_t \supset \mathcal{F}_t^N \ \forall t$.[68] Important additional information in this context includes the marks of the events other than the event type (for example, the size and direction of trades), data concerning news announcements that occur both 'overnight' and within normal trading hours, and data relating to stock exchange opening procedures.

A g-HawkesE(k) type intensity can be specified conditional on the marks of the events by making the jump that occurs in response to an event depend on that event's marks.[69] For example, consider the specification of the term $\tilde{\lambda}_{12}^{(j)}(t)$ in (24) and (25), and suppose that we are concerned with a bivariate model of trades and mid-quote events so that $\tilde{\lambda}_{12}^{(j)}(t)$ determines the ($j$th) effect of trades on the quote intensity. In (25), the jump in $\tilde{\lambda}_{12}^{(j)}(t)$ in response to a

---

[67]It is necessary to choose $T^{(m)}(\theta)$ carefully. For example, the procedure results in a test with extremely low power when the Box-Ljung statistic is used because there are almost always parameter values that make the sample variance very large indeed, both under the null and the alternative.

[68]Theorem 3.1 requires that $\mathcal{F}_t \supseteq \mathcal{F}_t^N \ \forall t$. Our diagnostic testing procedures therefore continue to be valid for models where we condition on a wider filtration than the internal history of the PP.

[69]Recall that we write a Marked Point Process as $\{T_i, Z_i\}_{i \in \{1,2..\}}$. The marks that we condition on here are (potentially) all of the elements of the vector $Z_i$ other than the event type. In this case an $\mathcal{F}_t$-intensity is specified where $\mathcal{F}_t$ is the internal history of $\{T_i, Z_i\}_{i \in \{1,2..\}}$.

trade is always equal to $\alpha_{12}^{(j)}$. The suggested extension is to set the jump that occurs in response to the $i$th trade equal to $Y_i$, where $Y_i$ is a parametrised function of the marks of that trade. Denoting the occurrence time and marks of the $i$th trade by $T_i^{(2)}$ and $Z_i^{(2)}$ respectively, the BV-g-HawkesE(k) type specification for $\tilde{\lambda}_{12}^{(j)}(t)$ thus becomes

$$\tilde{\lambda}_{12}^{(j)}(t) = \pi_{12}^{(j)}\tilde{\lambda}_{12}^{(j)}(\tau_{d-1})e^{-\rho_{12}^{(j)}(t-\tau_{d-1})} + \sum_{T_i^{(2)}\in[\tau_{d-1},t)} Y_i e^{-\beta_{12}^{(j)}(t-T_i^{(2)})}, \tag{38}$$

for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ...)$, where $Y_i = g(Z_i^{(2)})$ for some function $g$. Such an extension should prove useful in future market microstructure research.[70]

The effect of an overnight news announcement can be included by allowing the intensity at the start of the day to depend on the announcement data, with the effect damping down during the day in a manner analogous to the spillover effects in the g-HawkesE(k) models. Conditioning on data related to opening procedures could be performed in a similar manner. Finally, we note that if $Y(t)$ is a continuous time stochastic process with sample paths of locally bounded variation, and we take $\mathcal{F}_t = \sigma(N(s), Y(s) : 0 \leq s \leq t)$, then a possible generalisation of the univariate g-HawkesE(k) model is given by,[71]

$$\lambda(t) = \mu(t) + \sum_{j=1}^{k}\tilde{\lambda}_j(t) + \int_0^t f(t-s)dY(s), \tag{39}$$

where $\tilde{\lambda}_j(t)$ is as in (17), $f(.)$ is a deterministic function, and the integral is the stochastic Stieltjes integral with respect to $Y$.

This paper points to a number of challenges for future research. In terms of the specification of useful empirical models, the case where $M$ is moderately large (say, $M = 5$) and alternative forms of spillover effect such as the one discussed towards the end of Section 4.3 are of particular interest. An important theoretical challenge is the development of a body of asymptotic distribution theory following on from the work of Ogata (1978), including the extension of those results to the multivariate case and establishing results for the case where parameters are on the boundary of the parameter space.[72]

---

[70]If $Y_i$ is allowed to take negative values, then the question arises of how to ensure the positivity of the intensity. A possible solution (for BV-g-HawkesE(k) type models) is to set the non-deterministic component of the $m$th intensity equal to $\max\{0, \sum_{j=1}^{k}\tilde{\lambda}_{mm}^{(j)}(t) + \sum_{j=1}^{k}\tilde{\lambda}_{mn}^{(j)}(t)\}$.

[71]Ogata and Akaike (1982) make a similar observation.

[72]Ogata (1981) extends Lewis' thinning simulation algorithm to multivariate PPs which are absolutely con-

# 7   Conclusion

We develop here a continuous time econometric modelling framework for multivariate market event data in which the model is specified via the vector stochastic intensity. This is a powerful and natural approach to specifying multivariate models, since one considers how the intensity (conditional hazard) for each type of event changes as the information set is updated in continuous time. Using the specific example of generalised Hawkes models, a model structure is introduced in which the non-deterministic components of the intensity on trading day $d$ are specified recursively in terms of functionals of the intensity paths on previous days and the contributions of the events occurring on day $d$. This structure takes into account the existence of overnight periods when the stock market is closed and allows the dependence of the intensity on the events of previous trading days to be estimated. Analytic likelihoods are available for the models and diagnostic tests can be readily computed. Our empirical results highlight the importance of not imposing *a priori* assumptions such as a lack of dependence between trading days.

We show how to use a random change of time approach to construct diagnostic tests of parametric, multivariate PP models. This involves transforming the non-Poisson processes into Poisson processes and then exploiting the *i.i.d.* exponential property of the durations of the latter to construct the tests. The technique has widespread applicability since the restriction it imposes on the PP model is very natural in the context of financial market event data. Change of time techniques appear to have been exploited very little to date in the statistical literature on goodness of fit tests for counting processes. We propose various diagnostic tests, some of which are new to the econometrics PP literature, and the impact of parameter uncertainty on the testing procedures is considered.

A full bivariate PP model of the occurrence times of trades and changes to the mid-quote is presented for a NYSE stock. Importantly, the bivariate g-HawkesE(2) model allows the study of the two way interaction between trades and price changes. We find that not only do trades

---

tinuous with respect to the (suitably multivariate) standard Poisson process. This provides an efficient way to simulate from PPs that are specified via the stochastic intensity and should prove useful in future work, including the study of the finite sample properties of various statistics.

result in an increase in the intensity for mid-quote events (as is expected from the sequential trade models), but mid-quote events also result in increased trade intensity. The estimated mid-quote intensity is used to provide an approximation to the instantaneous price volatility of the stock and supports a central prediction of the Easley and O'Hara (1992) model – namely, that periods of low volatility tend to occur in periods when there are few trades. Finally, a comparison with previous related work is provided and suggested extensions to the modelling framework are discussed. An important strength is the ease with which the approach can be extended in order to model the evolution of the multivariate PP conditional on a wider filtration than its internal history.

# 8 Appendix

## 8.1 Data Adjustments

In all cases, the final data pertaining to trades and quotes was obtained as follows. First the times of all changes to the mid-quote were determined. Events (i.e. trades and quotes) occurring outside of normal trading hours (9:30 EST to 16:00 EST) were then deleted; for NYSE stocks the first reported trade of each calendar day was removed as this details the price and volume of the opening auction.[73] Trades that were later cancelled were also deleted.[74] The times of the events were then mapped onto $[0, \infty)$ using the data transformation described in Section 4.1, and the times of changes to the mid-quote were thinned to obtain the 'mid-quote events' that we model (see Section 5.2 for details). Events of the same type (i.e. two trades or two quotes) that have the same timestamp were then treated as a single event with that timestamp. In all of the datasets analysed here, the occurrence of such events was rare.[75]

The $M$-variate PPs that we deal with in this paper assign zero probability to the simultaneous occurrence of two market events (of either the same or different types). The treatment of two events of the same type with identical timestamps has already been discussed. The treatment of

---

[73]We thus exclude from our analyses data relating to the preopening period on NASDAQ and the opening auction of the NYSE.

[74]Only trade records with a TAQ correction indicator equal to 0 or 1 were included in the analysis. In the case of trades that underwent correction, the original occurrence time of the trade is used.

[75]E.g., in the case of the General Motors (GM) data analysed in Section 5.2, the number of trade events is reduced by only 0.26 per cent as a result of treating trades with identical timestamps as single trades; the reduction in the number of ('post thinning') mid-quote events was only 0.14 per cent.

the simultaneous occurrence of trades and mid-quote events is a more substantive issue. In the case of the GM data, approximately 11 per cent of the mid-quote events have exactly the same timestamp as a trade. We adjust the occurrence times of these mid-quote events as follows: if $x$ is the original occurrence time (in seconds), then the time becomes $x - 0.5 + U$ in the final dataset, where $U$ is the realisation of a uniform r.v. on $(0, 1)$.[76] This adjustment procedure is referred to below as Algorithm 1.

Some comments concerning this procedure are in order. First, although the reported timestamps (in whole seconds) of the mid-quote event and the trade are the same in the data, the actual occurrence times (in continuous time) will rarely be the same. By the actual occurrence time we mean the time of trade execution by the specialist and the time when the specialist made known his revised quotes to the trading crowd. Consider the situation where the specialist calls out in one sentence the details of a trade and new quotes set in response to that trade. These events might well receive the same timestamp in the data although in reality trade execution occurred first. An alternative to the approach adopted here would be to define a third type of event as occurring whenever a trade and a mid-quote event have identical timestamps. However, this would be to aggregate market outcomes that are quite different economically: such timestamps could be the result of a trade execution occurring just prior to a quote change; a quote change occurring just prior to a trade execution; or the events occurring further apart in real time but being reported with identical timestamps as a result of lengthier reporting delays.[77] We thus prefer to adopt the adjustment procedure described above and regard the simultaneous occurrence of a trade and a mid-quote event in the data as almost always the result of lags between the (non-simultaneous) actual occurrence times of the events in continuous time and the reported times.[78]

---

[76]If this procedure does not result in a time different to that of the trade, then a further draw from $U$ is generated. However, this was not necessary here.

[77]A number of studies highlight the reporting delays associated with this sort of data (see, for example, Hasbrouck, Sofianos, and Sosebee (1993) and Lee and Ready (1991)) and it is known that the reported times of the events and the actual occurrence times may differ by several seconds. The time delays tend to be (but are not always) greater for trades than for quotes as a result of the different reporting mechanism that is sometimes used for trades, and the delays also vary across events of the same type (see, for example, the discussion relating to trade reporting at p.9, Hasbrouck, Sofianos, and Sosebee (1993)).

[78]It has been suggested that the '5 second rule' of Lee and Ready (1991) be applied when modelling a bivariate system of trade and quote times. This involves delaying every quote time by five seconds. However, the results presented by Lee and Ready (1991) show a frequency distribution for the difference between the timestamp of the trade and the quote in circumstances where it is reasonable to believe that the actual occurrence times were very

Since it would be expected that the estimates of $\alpha_{mq}^{(1)}, \beta_{mq}^{(1)}, \alpha_{mq}^{(2)}, \beta_{mq}^{(2)}$ $(m = 1, 2; q \neq m)$ would be sensitive to the particular treatment of the identical timestamps that is adopted, we conducted a sensitivity analysis by comparing our results with those obtained using two other adjustment rules: in the first of these the adjusted mid-quote event time is given by $x + U$ (with $U$ defined as above), and in the second the trade with the identical timestamp is deleted from the data. The first (Algorithm 2) is designed to capture the assumption that the identical timestamps are very often the result of the specialist executing a trade and then very rapidly calling out details of the trade and the new quotes. The second rule (Algorithm 3) is an alternative way of avoiding strong *a priori* assumptions concerning the actual ordering of the mid-quote event and the trade. Since there are many more trades than quotes, Algorithm 3 results in only a small reduction in the number of trades. All three algorithms gave very similar MLEs for parameters other than $\alpha_{mq}^{(1)}, \beta_{mq}^{(1)}, \alpha_{mq}^{(2)}, \beta_{mq}^{(2)}$ $(m = 1, 2; q \neq m)$, and resulted in diagnostic tests that accepted at the 1% level with the exception of the $BL$ test for the trade intensity. The estimates of $(\alpha_{12}^{(1)}, \alpha_{12}^{(2)}, \beta_{12}^{(1)}, \beta_{12}^{(2)}; \alpha_{21}^{(1)}, \alpha_{21}^{(2)}, \beta_{21}^{(1)}, \beta_{21}^{(2)})$ were $(59.4, 6.3, 1163.5, 123.7; 81.6, 19.1, 436.1, 58.5)$ for Algorithm 2 and $(27.9, 2.04, 475.4, 55.4; 80.6, 19.7, 423.8, 57.5)$ for Algorithm 3. These should be compared with those given for Algorithm 1 in Table 4. It is the positive effect of mid-quote events on the trade intensity that is perhaps the more surprising of our findings. Crucially, although the estimates of $\alpha_{21}^{(1)}$ and $\alpha_{21}^{(2)}$ are smaller for Algorithms 2 and 3 than for Algorithm 1, they are still far from zero. For example, when Algorithm 3 is employed, imposing the restrictions $\alpha_{21}^{(j)} = 0$ $(j = 1, 2)$ on $\lambda_2(t)$ results in a reduction in the log-likelihood of 277 and in the excess dispersion test again rejecting at the 1% level. We conclude that our finding of a positive effect of mid-quote events on the trade intensity is not the result of having employed a data adjustment rule that was unduly biased in favour of finding such an effect.

## 8.2 Computation of MLEs

We describe in detail here the procedure used to compute MLEs for the univariate g-HawkesE($k$) models and discuss how to extend this procedure to the bivariate case. Evaluating the log-

---

close together in real time (and trading is slow). The distribution has a mode of zero and 38.3% of the trades are recorded *prior to* the quote change. This suggests that it is preferable not to adjust all quote times by a deterministic amount.

likelihood for the univariate g-HawkesE($k$) model given in (22) yields

$$
\begin{aligned}
l(\theta; T_1, ..., T_{N(T)}) &= -\int_0^T \mu(s\ ;\gamma)ds - \sum_{d=1}^{T/l}\sum_{j=1}^k \left\{ \pi_j/\rho_j(1 - e^{-l\rho_j})\tilde{\lambda}_{j,\theta}(\tau_{d-1}) + \right. \tag{40}\\
&\left. \sum_{\tau_{d-1}\leq T_i<\tau_d} \alpha_j/\beta_j(1 - e^{-\beta_j(\tau_d - T_i)}) \right\} + \sum_{d=1}^{T/l}\sum_{T_i\in A_d}\log\{\mu(T_i\ ;\gamma) + \\
&\sum_{j=1}^k \left[\pi_j\tilde{\lambda}_{j,\theta}(\tau_{d-1})e^{-\rho_j(T_i-\tau_{d-1})} + \alpha_j A_{d,j}(i)\right]\} + T,
\end{aligned}
$$

where $A_{d,j}(i) = \sum_{z:\tau_{d-1}\leq T_z<T_i} e^{-\beta_j(T_i-T_z)}$ for $d = 1, 2, ..., T/l$ (and we adopt the convention that an empty sum is equal to zero). When computing (40) in practice, we use the recursion in (17) to calculate the $\tilde{\lambda}_{j,\theta}(\tau_{d-1})$'s. We also make use of the following recursion in $i$

$$
A_{d,j}(i+1) = (1 + A_{d,j}(i))e^{-\beta_j(T_{i+1}-T_i)}, \tag{41}
$$

in order to improve computational efficiency. For the purposes of numerical optimisation, the log-likelihood was reparametrised in terms of $\log\alpha_1, ..., \log\alpha_k, \log\beta_1, \log(\delta_2/(1-\delta_2)), ..., \log(\delta_k/(1-\delta_k))$, $\log\pi_1, \log\rho_1, ..., \log\pi_k, \log\rho_k, \log\gamma_1, ..., \log\gamma_8$, where $\delta_j = \beta_j/\beta_{j-1}$ $(j = 2, ..., k)$. We thus impose the positivity constraints $\alpha_j > 0$, $\pi_j > 0, \rho_j > 0$ $(j = 1, ..., k)$, $\gamma_1 > 0, ..., \gamma_8 > 0$, and $\beta_1 > \beta_2 > ... > \beta_k > 0$.

The evaluation of the log-likelihood for the bivariate g-HawkesE($k$) model given by (26) and (27) is a straightforward extension of (40) above. The log-likelihood was reparametrised in the same way as for the univariate model (with the obvious extension to the bivariate case) when performing numerical optimisation. A slight complication arises in the need to calculate terms of the form

$$
A_{mq}^{(d,j)}(i) = \sum_{z:\tau_{d-1}\leq T_z^{(q)}<T_i^{(m)}} e^{-\beta_{mq}^{(j)}(T_i^{(m)}-T_z^{(q)})}, \tag{42}
$$

for all $T_i^{(m)} \in (\tau_{d-1}, \tau_d]$ (where $q = 2$ if $m = 1$ and $q = 1$ if $m = 2$). We then make use of the following recursion in $i$

$$
A_{mq}^{(d,j)}(i+1) = A_{mq}^{(d,j)}(i)e^{-\beta_{mq}^{(j)}(T_{i+1}^{(m)}-T_i^{(m)})} + \sum_{z:T_i^{(m)}\leq T_z^{(q)}<T_{i+1}^{(m)}} e^{-\beta_{mq}^{(j)}(T_{i+1}^{(m)}-T_z^{(q)})}. \tag{43}
$$

When estimating the univariate g-HawkesE(2) model, initial parameter values for the BFGS algorithm were obtained by estimating the following sequence of models:[79]

1) Non-homogeneous Poisson model given by $\lambda(t) = \mu_\gamma(t)$, where $\mu_\gamma(t)$ is given by (23);

2) HawkesE(1) model with $\bar{\gamma} = \hat{\gamma}$;

3) HawkesE(2) model with $\bar{\alpha}_1 = \bar{\alpha}_2 = 0.5\hat{\alpha}_1, \bar{\beta}_1 = \hat{\beta}_1, \bar{\beta}_2 = 0.99\hat{\beta}_2, \bar{\gamma} = \hat{\gamma}$;

4) g-HawkesE(2) model with $\bar{\alpha}_j = \hat{\alpha}_j, \bar{\beta}_j = \hat{\beta}_j, \bar{\pi}_j = 1, \bar{\rho}_j = \hat{\beta}_j, \bar{\gamma} = \hat{\gamma}$.

In order to estimate the bivariate g-HawkesE(2) model, we maximise $l_m(\theta_m) - T$ for $m = 1$ and $m = 2$ separately (where $l_m(\theta_m)$ is given by (27)). Initial parameter values were again obtained by maximising a sequence of log-likelihoods, commencing with the maximisation of $l_m(\theta_m) - T$ for the BV-HawkesE(2) model using as starting values the estimates obtained from the univariate HawkesE(2) model for $m$-type events in step 3) above:[80]

1) BV-HawkesE(2) model with

$$\bar{\alpha}_{mm}^{(j)} = \hat{\alpha}_j, \bar{\beta}_{mm}^{(j)} = \hat{\beta}_j, \bar{\alpha}_{mq}^{(j)} = 0, \bar{\beta}_{mq}^{(j)} = \hat{\beta}_j, \bar{\gamma}_m = \hat{\gamma}_m \, (q \neq m);$$

2) BV-g-HawkesE(2) model with

$$\bar{\alpha}_{mr}^{(j)} = \hat{\alpha}_{mr}^{(j)}, \bar{\beta}_{mr}^{(j)} = \hat{\beta}_{mr}^{(j)}, \bar{\pi}_{mr}^{(j)} = 1, \bar{\rho}_{mr}^{(j)} = \hat{\beta}_{mr}^{(j)}, \bar{\gamma}_m = \hat{\gamma}_m (\forall mr).$$

# References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics 6*, 701–726.

Aalen, O. O. and J. M. Hoem (1978). Random time changes for multivariate counting processes. *Scandinavian Actuarial Journal*, 81–101.

Akaike, H. (1977). *On entropy maximization principle*, pp. 27–41. Application of Statistics. Amsterdam: North-Holland.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica 69*, 683–734.

---

[79]Initial values of parameters are indicated by a bar, and a hat indicates a parameter estimate obtained from estimation of the model in the previous step.

[80]By the BV-HawkesE(2) model, we mean the BV-g-HawkesE(2) model with the restrictions ($\pi_{mq}^{(j)} = 1$ and $\rho_{mq}^{(j)} = \beta_{mq}^{(j)} \, \forall j, \forall mq$) imposed.

Arjas, E. (1986). *Stanford heart transplantation data revisited: a real time approach.* Modern Statistical Methods in Chronic Disease Epidemiology. New York: Wiley.

Arjas, E. and P. Haara (1988). A note on the exponentiality of total hazards before failure. *Journal of Multivariate Analysis 26*, 207–218.

Bauwens, L. and P. Giot (2001). *Econometric Modelling of Stock Market Intraday Activity*, Volume 38 of *Advanced Studies in Theoretical and Applied Econometrics.* Boston: Kluwer Academic.

Bowsher, C. G. (2002). Modelling security market events in continuous time: Intensity based, multivariate point process models. Economics Discussion Paper No. 2002-W22, Nuffield College, Oxford.

Brémaud, P. (1981). *Point Processes and Queues, Martingale Dynamics.* New York: Springer-Verlag.

Brémaud, P. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability 24*, 1563–1588.

Davis, R., T. Rydberg, and N. Shephard (2001). The Cbin model for counts: Testing common features in the speed of trading, quote changes, limit and market order arrivals. Mimeo.

Diebold, F., T. Gunther, and A. Tay (1998). Evaluating density forecasts, with applications to financial risk management. *International Economic Review 39*, 863–883.

Diebold, F., J. Hahn, and A. Tay (1999). Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange. *Review of Economics and Statistics 81*, 661–673.

Doornik, J. A. (2001). *Ox 3.0 - An Object-Oriented Matrix Programming Language.* London: Timberlake Consultants Ltd.

Dufour, A. and R. F. Engle (2000). Time and the price impact of a trade. *Journal of Finance 55*, 2467–2498.

Easley, D. and M. O'Hara (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics 19*, 69–90.

Easley, D. and M. O'Hara (1992). Time and the process of security price adjustment. *Journal of Finance 47*, 577–605.

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

Engle, R. F. and A. Lunde (2003). Trades and quotes: A bivariate point process. forthcoming in the Journal of Financial Econometrics.

Engle, R. F. and J. R. Russell (1997). Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of Empirical Finance 4*, 187–212.

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

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*, 71–100.

Grammig, J. and M. Wellner (2002). Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics 106*, 369–400.

Hamilton, J. D. and Ò. Jordà (2001). A model for the federal funds rate target. *Journal of Political Economy 110*, 1135–1167.

Harris, L. (1998). Optimal dynamic order submission strategies in some stylized trading problems. Monograph, forthcoming in Financial Markets, Institutions and Instruments 7(2).

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance 46*, 179–207.

Hasbrouck, J. (1996). *Modelling Market Microstructure Time Series*, Volume 14 of *Handbook of Statistics*, pp. 647–692. Amsterdam: Elsevier, North-Holland.

Hasbrouck, J. (1999). Trading fast and slow: Security market events in real time. Working Paper, Stern School of Business, New York University.

Hasbrouck, J., G. Sofianos, and D. Sosebee (1993). New York Stock Exchange systems and trading procedures. NYSE Working Paper 93-01.

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

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

Kamionka, T. (2001). La modélisation des données haute fréquence. Updated version of CREST Working Paper 2000-58.

Karr, A. F. (1991). *Point Processes and their Statistical Inference*. New York: Dekker.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Lee, C. M. C. and M. J. Ready (1991). Inferring trade direction from intraday data. *Journal of Finance 46*, 733–746.

Meyer, P. A. (1971). *Démonstration simplifiée d'un théorème de Knight*, Volume 190 of *Lecture Notes in Mathematics*. Berlin: Springer Verlag.

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

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory IT-27*, 23–31.

Ogata, Y. (1983). Likelihood analysis of point processes and its applications to seismological data. *Bulletin of the International Statistical Institute 50*, 943–961.

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

O'Hara, M. (1995). *Market Microstructure Theory*. Oxford: Blackwell.

Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math. 31, Part B*, 145–155.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics 23*, 470–472.

Russell, J. R. (1999). Econometric modelling of multivariate irregularly-spaced high-frequency data. Mimeo, University of Chicago, Graduate School of Business.

Russell, J. R. and R. F. Engle (1998). Econometric analysis of discrete-valued irregularly-spaced financial transactions data using a new autoregressive conditional multinomial model. University of California, San Diego Discussion Paper 98-10.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–131.

Smith, J. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting 4*, 283–291.

Spierdijk, L., T. E. Nijman, and A. H. van Soest (2002). Modeling comovements in trading intensities to distinguish sector and stock specific news. Working Paper, Tilburg University and CentER.

Thompson, S. B. (2001). Evaluating the goodness of fit of conditional distributions. Mimeo, Department of Economics, Harvard University.

Vere-Jones, D. and T. Ozaki (1982). Some examples of statistical estimation applied to earthquake data. *Ann. Inst. Statist. Math. 34*, 189–207.

Yashin, A. and E. Arjas (1988). A note on random intensities and conditional survival functions. *Journal of Applied Probability 25*, 630–635.