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.

30 January 2003

Abstract

A continuous time econometric modelling framework for *multivariate* financial 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 σ-field is updated continuously in time as new information arrives. The class of generalised Hawkes models is introduced which allows the estimation of the dependence of the intensity on the events of previous trading days. Analytic likelihoods are available and it is shown 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 New York Stock Exchange stock and the empirical findings are related to the theoretical and empirical market microstructure literature. The *two-way* interaction of trades and quote changes is found to be important empirically.

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

JEL classification: C32, C51, C52, G10.

1 Introduction

The availability of '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.[1] The aim of this paper is to develop a continuous time econometric modelling framework for *multivariate* market event data – that is, data that records the timing of several different *types* of market event. Many interesting economic questions concerning market microstructure can only be addressed using such multivariate models.

The main contributions of the paper may be summarised as follows. First, an intensity based approach to model specification is used to develop a new class of empirical models for multivariate market event data. The models are general enough to allow the issue of dependence between trading days to be

*Correspondence should be sent to clive.bowsher@nuffield.ox.ac.uk. Particular thanks are due to Neil Shephard, Bent Nielsen, David Cox and Frank Gerhard. Personal correspondence with Jeff Russell concerning the ACI model was very helpful and much appreciated. All computational work was performed using Version 3.0 of Ox (see Doornik(2001)). Financial support from the ESRC under awards R42200034061 and R00023839 is gratefully acknowledged.

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

addressed and analytic likelihoods are available for estimation. Second, the use of random changes of time to construct diagnostic tests for such multivariate models is established. A sufficient condition for the validity of the diagnostic testing procedures suggested previously by Russell (1999) is thus derived – a condition which is shown to be natural in the context of financial markets. Finally, applying the econometric methods developed in the paper to the trade and quote data of a New York Stock Exchange (NYSE) stock provides evidence that the *two-way* interaction between the timing of trades and quote changes is important empirically.

Econometrically speaking, multivariate market event data can be viewed as the realisation of a multivariate Point Process (PP): 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 occurrence time of the ith event and $Z_i \in \{1, 2, ..., M\}$ indicates the ith event's type. Whilst considerable progress has been made in modelling the univariate ($M = 1$) case using time series models of the time intervals or 'durations' between events (see, in particular, Engle and Russell (1997,1998)), multivariate extensions of this work have been slow to emerge in the econometrics literature.[2] I adopt a different approach in which the model is specified via the vector stochastic intensity. This provides a natural and powerful modelling framework for multivariate market event data. Each element of the stochastic intensity is a continuous time process that may be interpreted as the conditional hazard for the particular type of market event in question. The σ-field upon which the hazard is conditioned is updated continuously as new information arrives, thus allowing other types of event to influence the hazard as they occur in continuous time.[3] My approach is closest to that of Russell (1999), who also specifies a multivariate PP model via the stochastic intensity. A discussion of his modelling framework is provided towards the end of the paper.

A relevant issue to be considered is whether the obvious difficulties of modelling multivariate market event data could not be mitigated somewhat by adopting a simpler approach based on conventional time series methods (see Davis, Rydberg, and Shephard (2001)). Whilst such an approach has resulted in significant insights, the development of continuous time models for market event data is 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

[2]Univariate models of market event data, and Autoregressive Conditional Duration models in particular, are surveyed by Bauwens and Giot (2001).

[3]Hamilton and Jordà (2001) use an intensity based approach to model a discrete time PP and also note the advantage that this offers in terms of being able to incorporate immediately the effect of information that occurs within a duration.

of papers point to the economic significance of real, 'wall-clock' time (see, *inter alia,* Easley and O'Hara (1992), Hasbrouck (1999) and Dufour and Engle (2000)). Second, most potential practical applications such as volatility measurement and the design of optimal order submission strategies (see Harris (1998)) require that the models relate to real time. Finally, a standard time series analysis of aggregated data using fixed intervals of real time involves an undesirable loss of information since the characteristics and timing relations of individual events are lost.

The structure of the present paper is as follows. Section 2 briefly describes the aspects of PP theory that are central to the paper and discusses an intensity based approach to inference for PPs. Section 3 introduces a new class of models for financial market event data – the generalised Hawkes models. These allow the estimation of the nature of the dependence of the intensity on the events of previous trading days rather than imposing strong, *a priori* assumptions concerning this dependence. Section 4 builds on the suggestions of Russell (1999) for the construction of diagnostic tests for parametric, multivariate PP models. I show that the tests are best thought of as being based on the transformation of non-Poisson processes into standard Poisson processes using random changes of time. Various diagnostic tests are proposed, including some novel tests based on the superposition of the standard Poisson processes just mentioned. Section 5 then makes use of the models and diagnostic testing procedures developed earlier in order to analyse the two-way interaction between the timing of trades and mid-quote changes for a NYSE stock. The empirical findings are related to both the theoretical and empirical market microstructure literature. Section 6 explores connections between this work and previous research by Engle and Lunde (2003) and Russell (1999), 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 trade and quote dataset in order to deal with simultaneous events, and presents the results of a sensitivity analysis that establishes the robustness of the empirical findings to the particular adjustment rule that is employed; details of the numerical computation of maximum likelihood estimates for generalised Hawkes models are also provided.

2 Point Processes and Stochastic Intensities

This section reviews how a statistical model for multivariate PP data can be completely specified by a parametric family of stochastic intensities and establishes that a legitimate likelihood function and likelihood ratios are available. In so doing, definitions of an M-variate PP and its associated counting process, $N(t)$, are provided and some notation is established. Theorem 2.1 below 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 discussed. Textbook treatments of the martingale-based intensity theory of PPs are given in Brémaud (1981) and Karr (1991), from which I have drawn extensively in this section. Readers whose chief interest is in the new models and their empirical application rather than in the theoretical underpinning of the econometric methods may prefer to proceed to Section 3. They should, however, note Section 2.1 and the interpretation of the intensity as a conditional hazard in equation (4).

2.1 Point and counting processes

As was noted previously, multivariate market event data can be viewed as the realisation of an M-variate PP. I shall use T_i to denote the random occurrence time of the ith event and Z_i to denote the ith event's type.

Definition 1 ***Point Process.*** *Let* $\{T_i\}_{i\in\{0,1,2,\ldots\}}$ *be a sequence of nonnegative random variables on some probability space* $(\Omega, \mathcal{F}, P)$ *such that* $T_0 = 0$ *and* $0 < T_i \leq T_{i+1}$ *for* $i \geq 1$. *Then the sequence* $\{T_i\}$ *is called a point process on* $[0, \infty)$. *The point process is simple if and only if* $T_i < T_{i+1}$ $\forall i$.

Definition 2 ***M-variate point process and counting process.*** *Let* $\{T_i\}_{i\in\{1,2,\ldots\}}$ *be a simple point process on* $[0,\infty)$, *defined on* $(\Omega, \mathcal{F}, P)$, *and let* $\{Z_i\}_{i\in\{1,2,\ldots\}}$ *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* $\{T_i, Z_i\}$.[4]

[4] There is a one-one mapping between the M-variate point and counting processes for all $M \geq 1$. It is therefore permissible to refer to 'the M-variate PP, $N(t)$'.

Note that the definition of an M-variate PP rules out the possibility of the simultaneous occurrence of two events (of either the same or different types) since the PP $\{T_i\}$ is taken to be simple.[5]

Some comments on notation and terminology now follow. The ith event is said be of 'type m' for a particular realisation if and only if $Z_i(\omega) = m$ (where $m \in \{1, 2, ..., M\}$). I shall call $N_m(t)$, the process which counts the number of events of type m that have occurred up to and including time t, the mth marginal counting process, and will use $T_i^{(m)}$ to denote the occurrence time of the ith type m event.[6] 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,

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

The *internal history* (or natural filtration) 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 σ-field of all subsets of $\{1, 2, ..., M\}$. Notice that in the univariate case ($M = 1$) the definitions simplify to give $N(t) = \sum_{i \geq 1} 1(T_i \leq t)$ and $\mathcal{F}_t^N = \sigma(N(s) : 0 \leq s \leq t)$.

2.2 Stochastic intensities and likelihoods

The distinguishing feature of my approach is that the M-variate PP model is specified via the vector stochastic intensity, $\lambda(t) = (\lambda_m(t))_{m=1}^M$. This perspective is very natural and productive if one is concerned with developing new models – one considers how the intensity for type m events, $\lambda_m(t)$, changes in continuous time as new 'information' arrives. As will be shown below, the intensity can be interpreted as a conditional hazard function. A formal definition of the stochastic intensity of a simple PP is as follows.

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

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

for all t, s such that $0 \leq t \leq s$, then $h(t)$ is the $(P, \mathcal{F}_t)$-intensity of $N(t)$.[7]

[5] When $M > 1$, I shall 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.

[6] That is, $\{T_i^{(m)}\}_i$ is the simple PP that has $N_m(t)$ as its counting process. Note that $T_0^{(m)} := 0$.

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

In the multivariate case ($M > 1$), the M-variate process $\lambda(t) = (\lambda_m(t))_{m=1}^M$ is 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 $N_m(t)$ for $m = 1, ..., M$.[8] This paper will usually be concerned with the case $\mathcal{F}_t = \mathcal{F}_t^N$.

Suppose now that $N_m(t)$ is observed on $[0, T]$ and has the $(P, \mathcal{F}_t)$-intensity $\lambda_m(t)$, that the sample paths of $\lambda_m(t)$ are left-continuous with right-hand limits and that $\{\lambda_m(t)\}_{0 \le t \le T}$ is bounded by an integrable random variable (hereafter r.v.). Define $\lambda_m(t+) := \lim_{s \downarrow t} \lambda_m(s)$. Then $\lambda_m(t+)$ can be interpreted as a conditional hazard function since

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

(see, for example, Aalen (1978, Lemma 3.3)).[9]

The following theorem is of fundamental importance since it establishes that legitimate likelihood functions and ratios are available for the new models.

Theorem 2.1 *Let $(N_m(t))_{m=1}^M$ be an M-variate point process on $[0, T]$, $0 < T < \infty$, 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*

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

and define the probability measure P by the following density with respect to P_0

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

Then $N_m(t)$ has $(P, \mathcal{F}_t^N)$-intensity $\lambda_m(t)$ (for $m = 1, ..., M$).[10] 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. Apply the proof of Karr (1991, Theorem 5.2) setting the baseline intensity equal to $(1, 1, ..., 1) = \mathbf{1}_M$. For the uniqueness of P, see Brémaud (1981, Theorem T8, p.64). ■

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

[9] Intuitively, it is the right continuous process $\lambda_m(t+)$ rather than $\lambda_m(t)$ that appears in 4, because if there is a jump at time t then $\int_t^{t+dt} \lambda_m(u) du \simeq \lambda_m(t+) dt \neq \lambda_m(t) dt$.

[10] 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.

2.3 Intensity based inference

In this paper each statistical 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_θ, such that $N(t)$ has $(P_\theta, \mathcal{F}_t^N)$-intensity $\lambda_\theta(t)$. The implication is that one need only specify $\{\lambda_\theta(t)\}_{\theta\in\Theta}$ in order to specify the statistical model (i.e. the set of potential Data Generating Processes under consideration), $\{P_\theta\}_{\theta\in\Theta}$. 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.

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 simple null hypothesis possesses the standard χ^2 asymptotic null distribution. Since the focus is on stationary PPs, Ogata (1978) considers a PP on $(-\infty, +\infty)$, $\{T_i;\, i = 0, \pm 1, \pm 2, ...\}$, which is observed during the interval $[0, T]$. The asymptotic covariance matrix is shown to be $I(\theta_0)^{-1}$, where θ_0 is the true parameter value (assumed to be in the interior of the parameter space), $I(\theta_0)$ is the matrix $\{I_{ij}(\theta_0)\}_{ij}$ and

$$I_{ij}(\theta_0) = E_{\theta_0}\left[\frac{1}{\lambda^*_{\theta_0}(t)}\frac{\partial\lambda^*_{\theta_0}(t)}{\partial\theta_i}\frac{\partial\lambda^*_{\theta_0}(t)}{\partial\theta_j}\right] = -E_{\theta_0}\left[\frac{1}{T}\frac{\partial^2 l^*_T(\theta_0)}{\partial\theta_i\partial\theta_j}\right]. \tag{7}$$

In equation (7), $\lambda^*_{\theta_0}(t)$ denotes the *complete intensity* and $l^*_T(\theta_0)$ the (theoretical) conditional log-likelihood under the information from the infinite past. $l^*_T(\theta_0)$ is given by the logarithm of (6) with $\lambda(t)$ replaced by $\lambda^*(t)$ and $M = 1$. Under the assumed conditions, it is also shown that

$$-E_{\theta_0}\left[(1/T)(\partial^2 l_T(\theta_0)/\partial\theta_i\partial\theta_j)\right] \to I_{ij}(\theta_0) \quad \forall i, j, \tag{8}$$

where $l_T(\theta_0)$ is the exact log-likelihood for the statistical model. Thus the standard asymptotic distribution for a MLE applies, but with an additional subtlety introduced: it is $l^*_T(\theta_0)$ rather than $l_T(\theta_0)$ that appears in (7). Nevertheless, the convergence result in (8) motivates the use of the normed Hessian matrix, $(1/T)(\partial^2 l_T(\hat{\theta}_T)/\partial\theta\partial\theta')$, to estimate $\{I_{ij}(\theta_0)\}_{ij}$. Only a flavour of the results of Ogata (1978) has been given here and further details may be found in Bowsher (2002). The development of asymptotic distribution theory following on from Ogata's work is an important area for future research. Results for the multivariate and non-stationary cases would be of particular interest in the context of financial market event data.

Thus far I have advocated an intensity based approach to the specification of parametric, multivariate PP models and have shown that ML methods are available for estimation. The development of diagnostic testing procedures for such models using random changes of time is one of the contributions of this paper, and will be discussed in Section 4 after having first introduced the new generalised Hawkes models. The aim at this stage is to fix ideas and to provide the reader with a concrete example of a stochastic intensity, before returning to a higher level of generality for the discussion of diagnostic testing.

3 Generalised Hawkes Models

This section will work towards a definition of the bivariate generalised Hawkes (g-Hawkes) model that will later be used to analyse the trade and quote dataset in Section 5. The concepts introduced here can readily be extended to the general multivariate case ($M \geq 2$). The bivariate model has two important features. Rather obviously, the intensity for type 1 events depends on the history of type 2 events and vice versa. In addition, the models are general enough to allow the estimation of the dependence of the intensity on the events of previous trading days, thus taking into account the existence of overnight periods when the stock market is closed. In order to clarify the presentation that follows, it is useful first to discuss the data transformation that is used throughout the paper.

3.1 Data transformation

Since equity markets do not operate continuously, the researcher is faced with the question of how to model data generated during trading days that are separated by overnight periods.[11] The following approach is adopted here. Time zero is taken to be 09:30 on the first trading day of the dataset and the data pertaining to each of the 6.5 hour trading 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 09:30 of an event occurring on the dth trading day included in the dataset ($d = 1, 2, ...$), then that event appears as an event at time $x + 6.5(d-1)$ in the final dataset. Treating the data as a realisation of a single PP on $[0, T]$ then allows the straightforward use of existing theorems in the PP literature. However, it is important to take into account the overnight periods when the market is closed and to model carefully how the intensity depends on events that occurred on previous trading

[11] The normal opening hours of the NYSE are the 6.5 hour period between 09:30 EST and 16:00 EST (Eastern Standard Time).

days in real time. The generalised Hawkes models introduced below allow the nature of this dependence to be estimated from the data.[12]

3.2 Generalised HawkesE(k) models

In order to simplify the exposition, the univariate generalised Hawkes process will be described before then extending the discussion to the bivariate case. The intensity of the univariate process is defined recursively in terms of the levels of the stochastic 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 data transformation described above, the real half-line is partitioned into intervals of length l corresponding 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 $\tau_d = l \cdot d$ $(d = 0,1,2,...)$.[13]

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

$$\lambda(t) = \mu(t) + \sum_{j=1}^{k} \tilde{\lambda}_j(t), \tag{9}$$

where $\mu(t)$ is a positive, deterministic function, $\tilde{\lambda}_j(0) = 0$, and

$$\tilde{\lambda}_j(t) = \pi_j \tilde{\lambda}_j(\tau_{d-1}) e^{-\rho_j(t-\tau_{d-1})} + \int_{[\tau_{d-1},t)} \alpha_j e^{-\beta_j(t-u)} dN(u), \tag{10}$$

for $\tau_{d-1} < t \leq \tau_d$ $(d = 1,2,....)$, where $\alpha_j \geq 0, \beta_j > 0, \pi_j \geq 0$, and $\rho_j > 0$.

The stochastic intensity of the g-HawkesE(k) model is thus the sum of a deterministic component, $\mu(t)$, and k stochastic components, $(\tilde{\lambda}_j(t))_{j=1}^{k}$.[14] Equation (10) 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 jth component at the end of the previous trading day, and the contributions of events occurring prior to time t on day d. I refer to the first term in (10) as the (jth) intensity 'spillover effect' between trading days. The purpose of

[12]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.

[13]$l = 6.5$ in the empirical section of the paper.

[14]The 'E(k)' nomenclature refers to the superposition of k stochastic 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.

including such terms in the model specification is discussed below. Evaluating the second term in (10) yields $\sum_{i:\tau_{d-1}\leq T_i<t}\alpha_j e^{-\beta_j(t-T_i)}$. Its sample paths are left-continuous, jumping up by an amount α_j in response to the occurrence of an event and then decaying exponentially (according to the parameter β_j) until the occurrence of the next event. The univariate g-HawkesE(k) model can readily be extended to the bivariate case by including terms that capture the effect of type 2 events on the intensity for type 1 events, and vice versa.

Definition 5 ***Bivariate (BV) g-HawkesE(k) model.*** *The model is defined by the vector stochastic intensity* $(\lambda_1(t), \lambda_2(t))'$*, where*

$$\lambda_m(t) = \mu_m(t) + \sum_{j=1}^{k} \tilde{\lambda}_{mm}^{(j)}(t) + \sum_{j=1}^{k} \tilde{\lambda}_{mq}^{(j)}(t), \tag{11}$$

for $m = 1, 2$*, with* $\mu_m(t)$ *a positive, deterministic function,* $q = 2$ *if* $m = 1$ *and* $q = 1$ *if* $m = 2$*, and*

$$\tilde{\lambda}_{mr}^{(j)}(t) = \pi_{mr}^{(j)}\tilde{\lambda}_{mr}^{(j)}(\tau_{d-1})e^{-\rho_{mr}^{(j)}(t-\tau_{d-1})} + \int\limits_{[\tau_{d-1},t)} \alpha_{mr}^{(j)} e^{-\beta_{mr}^{(j)}(t-u)} dN_r(u), \tag{12}$$

for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ...)$, $\tilde{\lambda}_{mr}^{(j)}(0) = 0$*, where* $mr \in \{1, 2\} \times \{1, 2\}$*. The parameter restrictions* $\alpha_{mr}^{(j)} \geq 0, \beta_{mr}^{(j)} > 0, \pi_{mr}^{(j)} \geq 0, \rho_{mr}^{(j)} > 0$ $(\forall mr$ *and* $\forall j)$ *apply.*

Note the presence of the terms $\tilde{\lambda}_{mq}^{(j)}(t)$ in (11), which allow the occurrence of type q events to influence the intensity for type m events.[15] The essential building block of the model has not changed in moving from the univariate to the bivariate case, as is evident from a comparison of equations (10) and (12). The BV g-HawkesE(k) model nests the important case where there is no dependence between trading days. This occurs when $\pi_{mr}^{(j)} = 0$ ($\forall mr$ and $\forall j$) since there are then no intensity spillover effects between days for either $\lambda_1(t)$ or $\lambda_2(t)$. The BV g-HawkesE(k) process is also equivalent to the 'mutually exciting' process of Hawkes (1971) when the restrictions ($\pi_{mr}^{(j)} = 1$, $\rho_{mr}^{(j)} = \beta_{mr}^{(j)}$ $\forall mr$ and $\forall j$) are imposed.[16] Note that this implies a very restrictive form for the spillover effects. The Hawkes (1971) model – referred to here as the BV HawkesE(k) model – can be written as

$$\lambda_m(t) = \mu_m(t) + \sum_{r=1}^{2}\sum_{j=1}^{k} \int\limits_{(0,t)} \alpha_{mr}^{(j)} \mathrm{e}^{-\beta_{mr}^{(j)}(t-u)} dN_r(u), \tag{13}$$

[15] By definition, $q \neq m$. Note that both $\tilde{\lambda}_{mm}^{(j)}(t)$ and $\tilde{\lambda}_{mq}^{(j)}(t)$ are defined by equation (12).

[16] 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), Ozaki (1979), and Ogata and Akaike (1982); and application of such models to earthquake data is treated by Vere-Jones and Ozaki (1982) and Ogata (1983).

for $m = 1, 2$, where $\mu_m(t) > 0$, and $\alpha_{mr}^{(j)} \geq 0, \beta_{mr}^{(j)} > 0$. The integral in (12) is the same as that in (13) except that the range of the former is restricted to $[\tau_{d-1}, t)$. Hence the name generalised Hawkes models.

There are two basic motivations for adopting a recursive model specification that incorporates intensity spillover effects from one trading day to the next. First, the generalised Hawkes models allow the nature of the dependence between trading days to be estimated from the data, rather than imposing untestable, *a priori* assumptions concerning this dependence. Two types of assumption have been adopted in previous work – either the data is treated as the realisations of independent PPs (each PP corresponding to a trading day) or is viewed as the realisation of a single PP on $[0, T]$ (after the removal of the overnight periods), without taking the special nature of the times $(\tau_1, \tau_2, ...)$ that correspond to the ends of the trading days into account. (The second of these approaches would be exemplified by fitting the Hawkes (1971) model to the data rather than using the generalised HawkesE(k) model.) By adopting a flexible specification, the g-HawkesE(k) models nest both of these approaches whilst also allowing for more general spillover effects. In practice, datasets have been encountered in this work for which there are significant spillover effects and for which there appears to be no dependence between trading days.[17] The g-HawkesE(k) models also make it possible in principle to formally test the important hypothesis that there is no dependence between trading days. The second motivation for adopting a recursive model specification is that the approach can easily be extended in order to condition on additional information such as an overnight news announcement or a stock exchange opening procedure that occurs 'between' trading days. Another additive term depending on the additional data would enter (12), with the effect damping down during the trading day in a manner analogous to the spillover effects.

The model structure introduced here in which the stochastic components of the intensity on trading day d are specified recursively in terms of functionals of the paths of those components on previous days and the contributions of the events occurring on day d is very general, and provides a useful framework for approaching the issue of dependence between trading days in PP models of financial markets. It would be interesting in future work to explore alternative specifications of the spillover effect in generalised Hawkes models, for example, a specification in which the spillover effect depends on the entire path of the component of the intensity during the previous day via the term $\int_{\tau_{d-2}}^{\tau_{d-1}} W(\tau_{d-1} - s)\tilde{\lambda}_{mr}^{(j)}(s)ds$, where $W(.) \geq 0$ is a non-negative 'weighting' function.

[17] For an example of the former, see the results of fitting the univariate g-HawkesE(2) model to the times of trades of a NASDAQ stock in Bowsher (2002, Table 2); for the latter, see Section 5 of this paper.

3.3 Hypothesis testing and computation of MLEs

I conclude this discussion of generalised Hawkes models by briefly considering the computation of MLEs and hypothesis testing in the context of the BV g-HawkesE(k) model. Taking the logarithm of equation (6) yields the following log-likelihood for the BV-g-HawkesE(k) model

$$l(\theta) = l_1(\theta_1) + l_2(\theta_2), \tag{14}$$

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

$$l_m(\theta_m) = \sum_{d=1}^{T/l} \left\{ \int_{A_d} (1 - \lambda_m(s;\theta_m))ds + \int_{A_d} \log \lambda_m(s;\theta_m) dN_m(s) \right\}, \tag{15}$$

where $\lambda_m(s;\theta_m)$ is given by (11) and (12), $A_d = (\tau_{d-1}, \tau_d]$, and $l_m(\theta_m)$ has been decomposed into the contributions of the different trading days. This decomposition allows the use of the recursive specification given by (12) in order to compute the log-likelihood more efficiently. In the case of the g-HawkesE(k) models, integration of the sample path of the intensity (with respect to Lebesgue measure) can be performed analytically. Evaluating (15) thus yields

$$\begin{aligned} l_m(\theta_m) \quad = \quad & T - \int_0^T \mu_m(s;\theta_m)ds + \sum_{d=1}^{T/l} \sum_{T_i^{(m)} \in A_d} \{\log \lambda_m(T_i^{(m)};\theta_m)\} \\ & - \sum_{s=1}^{2} \sum_{d=1}^{T/l} \sum_{j=1}^{k} \Big\{ \pi_{ms}^{(j)}/\rho_{ms}^{(j)} (1 - \mathrm{e}^{-l\rho_{ms}^{(j)}}) \tilde{\lambda}_{ms}^{(j)}(\tau_{d-1};\theta_m) + \\ & \sum_{\tau_{d-1} \le T_i^{(s)} < \tau_d} \alpha_{ms}^{(j)}/\beta_{ms}^{(j)} (1 - \mathrm{e}^{-\beta_{ms}^{(j)}(\tau_d - T_i^{(s)})}) \Big\}. \end{aligned} \tag{16}$$

Numerical optimisation of $l_m(\theta_m) - T$ was performed using the MaxBFGS algorithm with numerical derivatives (see Doornik (2001)). In order to estimate the BV 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. 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 given in the Appendix.

Consider now testing the null $H_0 : \alpha_{mr}^{(j)} = 0$, against the alternative $H_1 : \alpha_{mr}^{(j)} > 0$ (the maintained hypothesis being that $\alpha_{mr}^{(j)} \ge 0$), where $mr \in \{1,2\} \times \{1,2\}$ as in (12). This testing problem violates two standard regularity conditions. The parameter value under the null lies on the boundary of the maintained hypothesis, and there are nuisance parameters $(\pi_{mr}^{(j)}, \rho_{mr}^{(j)}, \beta_{mr}^{(j)})$ that are identified under the alternative but not under the null. The same comments apply to a test of $H_0 : \pi^{(j)}_{mr} = 0$ against the alternative $H_1 : \pi^{(j)}_{mr} > 0$ (in which case $\rho^{(j)}_{mr}$ is unidentified under the null). A consequence is that the LR tests of the various hypotheses will not possess the standard χ^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). 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.

4 Diagnostic Testing

Before we proceed to some empirical work using the bivariate g-HawkesE(k) model, we consider the question of diagnostic testing. The setting will again be that of a general parametric, M-variate PP model specified by the family of $\mathcal{F}^N_t$-intensities $\{\lambda(t;\theta)\}_{\theta\in\Theta}$, where $\lambda(t;\theta) = (\lambda_m(t;\theta_m))^M_{m=1}$ is a positive, $\mathcal{F}^N_t$-predictable process for all $\theta \in \Theta$ (see Section 2.3). The issue is how to construct diagnostic tests for such a model and what conditions any suggested procedure imposes on the PP model in order to ensure its validity. Building on the work of Russell (1999), I show that diagnostic testing may be performed by transforming the non-Poisson processes, $N_m(t)$, into standard Poisson processes using random changes of time. A proof of the validity of the diagnostic testing procedures is given that imposes a natural condition on the PP model, thus establishing their widespread applicability. Section 4.2 proposes various diagnostic tests, some of which are new to the econometrics literature. In addition, two feasible procedures that correctly take account of the parameter uncertainty are discussed.

Consider the sequences $\{e^{(m)}_i(\theta^*_m)\}_i$ for $m = 1, ..., M$, where $e^{(m)}_i(\theta^*_m) := \int_{T^{(m)}_i}^{T^{(m)}_{i+1}} \lambda_m(s;\theta^*_m)ds$ ($i = 0, 1, ...$), θ^*_m denotes the true parameter vector, the integrand is the intensity for type m events and $(T^{(m)}_i, T^{(m)}_{i+1}]$ is the duration between adjacent type m events. Russell (1999) suggests the use of these sequences as the basis for diagnostic testing but does not give a proof of the *i.i.d.* exponential property of each sequence. Rather, the paper indicates that the results of Yashin and Arjas (1988) concerning the 'exponential formula' might be used to establish this property. However, this requires that $F^{(m)}_i(x) = \Pr[S^{(m)}_i \leq x|\mathcal{G}_x]$ is absolutely continuous (with respect to Lebesgue measure), where $S^{(m)}_i := T^{(m)}_{i+1} - T^{(m)}_i$ and $\mathcal{G}_x$ is the observed history on which the assessment of the hazard is based.[18] It is important to note

[18] A proof based on the results of Yashin and Arjas (1988) is not a trivial matter and the issues involved are not explored in any depth here. For example, $\mathcal{G}_x$ is not explicitly defined. My purpose is to highlight the restrictive absolute continuity

that if one seeks to apply the exponential formula to establish the property in question, the conditioning here has to be 'dynamic' (i.e. $\mathcal{G}_x$ must depend on x). This is because events of other types, not equal to m, can occur within the duration and λ_m is conditional upon this extra information. The absolute continuity condition is undesirably restrictive in this context since it does not allow $F_i^{(m)}$ to exhibit jumps in response to the occurrence of type q events ($q \neq m$). There follows a general proof of the *i.i.d.* exponential property of the $\{e_i^{(m)}(\theta_m^*)\}_i$ sequences that imposes a much weaker condition on the PP model.

4.1 A random change of time argument

I shall use a random change of time argument to identify each $\{e_i^{(m)}(\theta_m^*)\}_i$ sequence as the durations of a standard Poisson process obtained by transforming the PP $N_m(t)$, thus establishing the *i.i.d.* exponential property of each sequence. Theorem 4.1 below (derived from Brémaud (1981, Theorem T16, p.41)) forms the basis for the proof of the result in Corollary 4.2. The main contribution of this section is in establishing the use of change of time techniques in the construction of diagnostic tests for parametric PP models. Using these techniques, it is then possible to isolate (17) as a sufficient condition for the validity of the diagnostic testing procedures suggested by Russell (1999). Known results in the change of time literature also provide a means of combining the information in the M sequences $[\{e_i^{(m)}(\theta_m^*)\}_i; m = 1, ..., M]$. Random changes of time appear to have been exploited surprisingly little to construct goodness-of-fit tests (for point and counting processes) in the statistical literature.[19]

Theorem 4.1 *Let $N(t)$ be an M-variate point process on $[0, \infty)$ with internal history $\mathcal{F}_t^N$ and $M \geq 1$. Also 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 $(P_\theta, \mathcal{F}_t)$-intensity $\lambda_m(t; \theta_m)$, where $\lambda_m(t; \theta_m)$ satisfies*

$$\int_0^\infty \lambda_m(t; \theta_m) dt = \infty \quad P_\theta\text{-}a.s. \tag{17}$$

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

$$\int_0^{\tau_m(t)} \lambda_m(s; \theta_m) ds = t. \tag{18}$$

condition that such a proof would necessarily involve. Further discussion is given in Bowsher (2002, p.41).

[19] Exceptions of which I am 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.

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

$$\tilde{N}_m(t) = N_m(\tau_m(t)) \tag{19}$$

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

Proof. Apply the proof of Brémaud (1981, Theorem T16, p.41)) 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 a change of time is applied to each marginal counting process, $N_m(t)$, is straightforward since $N_m(t)$ is itself a univariate PP.[20] Under certain additional assumptions, Aalen and Hoem (1978) also prove that $\tilde{N}_1(t), ..., \tilde{N}_M(t)$ are independent Poisson processes.[21] It is now possible to state and prove the main result of this section.

Corollary 4.2 *Let $\{T^{(m)}_i\}_{i\in\{0,1,2..\}}$ $(m = 1, ..., M)$ be the point process that has the counting process $N_m(t)$ referred to in Theorem 4.1. Also define $e^{(m)}_i(\theta_m) := \int_{T^{(m)}_i}^{T^{(m)}_{i+1}} \lambda_m(s; \theta_m)ds$ $(i = 0, 1, 2, ...)$. Then, under P_θ, $\{e^{(m)}_i(\theta_m)\}_{i\in\{0,1,2...\}}$ is an i.i.d. sequence of Exponential random variables with mean 1 for $m = 1, ..., M$.*

Proof. Let $\{\tilde{T}^{(m)}_i\}_{i\in\{0,1,2..\}}$be the point process that has the counting process $\tilde{N}_m(t)$ referred to in Theorem 4.1 and let s satisfy $\tilde{T}^{(m)}_{i-1} \leq s < \tilde{T}^{(m)}_i$ (for some $i \geq 1$). Then (19) implies that $\tau_m(s) < T^{(m)}_i \leq \tau_m(\tilde{T}^{(m)}_i)$. Since τ_m is continuous, letting $s \uparrow \tilde{T}^{(m)}_i$ establishes that $\tau_m(\tilde{T}^{(m)}_i) = T^{(m)}_i$. It then follows from (18) that $\tilde{T}^{(m)}_{i+1} - \tilde{T}^{(m)}_i = \int_{T^{(m)}_i}^{T^{(m)}_{i+1}} \lambda_m(s; \theta_m)ds$ for $i = 0, 1, 2, ...$ The result is now immediate since the $e^{(m)}_i(\theta_m)$'s are the durations of a standard Poisson process. ■

Note that the only condition required for this result is given in (17) and that this holds if and only if $\lim_{t\uparrow\infty} N_m(t) = \infty$ P_θ-a.s. (see Brémaud (1981, Lemma L17, p.41)). This condition is 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 type m events ever occur after some point in time. A sufficient condition

[20] Theorem T16 of Brémaud (1981) does not require the intensity to be with respect to the *internal* history of the univariate PP in question ($\mathcal{F}^N_{m,t}$ 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 $\mathcal{F}^N_{m,t} \subset \mathcal{F}^N_t \subseteq \mathcal{F}_t$ $\forall t$.

[21] See Section 4.5 of Aalen and Hoem (1978); they require that $\mathcal{F}_t = \mathcal{F}^N_t$ and that each 'distribution function' $\Pr[T_{i+1} - T_i \leq x, Z_{i+1} = m|\mathcal{F}^N_{T_i}]$ 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 $\mathcal{F}^N_t \subset \mathcal{F}_t$ $\forall t$.

for the BV g-HawkesE(k) model to satisfy (17) for $m = 1, 2$, is that

$$\mu_m(t) \geq \epsilon_m > 0 \ \forall t \text{ and } m = 1, 2, \tag{20}$$

where ϵ_m is some positive constant, since this implies that $\int_0^\infty \mu_m(t)dt = \infty$ $(m = 1, 2)$. The BV g-HawkesE(k) model used to analyse the trade and quote dataset in Section 5 has $\mu_m(t)$ given by equation (23), which clearly satisfies (20).

It is worth noting that whilst it is often the case for univariate models ($M = 1$) that the $\{e_i(\theta)\}$ series can be obtained by transforming to exponential r.v.'s the particular *i.i.d.* uniform series used for diagnostic evaluation by authors such as Smith (1985), Shephard (1994), Kim, Shephard, and Chib (1998) and Diebold, Gunther, and Tay (1998), the diagnostic procedures for multivariate PPs developed here cannot be interpreted as a multivariate extension of this work.

4.2 Constructing diagnostic tests

The aim is to test whether there exists a $\theta^* \in \Theta$ such that the Data Generating Process (DGP) is P_{θ^*}. In order to develop an approach to this problem, I first discuss the behaviour of test statistics when the DGP is known to be P_{θ^*} before considering possible solutions to the parameter uncertainty (or 'nuisance parameter') problem. Corollary 4.2 states that if the DGP is P_{θ^*}, then $e_i^{(m)}(\theta_m^*) \sim$ *i.i.d.* $Exp(1)$ $(m = 1, ..., M)$. There are many possibilities for the construction of diagnostic test statistics based on the M series $[\{e_i^{(m)}(\theta_m^*)\}_i; m = 1, ..., M]$. Consider first tests that are based, say, on the mth 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 test is the test of no excess dispersion suggested by Engle and 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 $e_i^{(m)}(\theta_m^*)$'s. The test statistic has an asymptotic $\mathcal{N}(0, 1)$ distribution when $\{e_i^{(m)}(\theta_m^*)\}$ is *i.i.d.* $Exp(1)$. Alternatively, the series can be transformed to yield an *i.i.d.* uniform series. 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 a series is *i.i.d.* uniform and explains how to compute quantiles of the (finite sample) null distributions by simulation.

Other diagnostic testing possibilities arise by combining the M series in some way, either prior to or after the calculation of the test statistics. It was noted above that Aalen and Hoem (1978) have proved under quite general conditions that 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. Additional diagnostic tests may therefore be based on the PP formed by 'pooling' the points associated with the M $\{e_i^{(m)}(\theta_m^*)\}$ series (i.e. sorting the collection of points $\{\tilde{T}_i^{(m)}\}_{i,m}$ into ascending order). Tests of the *i.i.d.* $Exp(1)$ property of the durations between those points (after rescaling the durations by multiplying by M) may then be calculated as described above.[22] Such tests have not been suggested before in the literature. The possibility of either combining multiple tests based on the mth series into a single test or combining tests based on different series has not been pursued here. The distributional results of Thompson (2001) can be used to do the former. Two tests based on, say, the mth and qth ($m \neq q$) series will be independent under the conditions of Aalen and Hoem (1978).

In the empirical section of the paper, 3 diagnostic test statistics are reported for each m ($m = 1, 2$): the Box-Ljung (BL) tests that the first 15 autocorrelations are all equal to zero for the $\{e_i^{(m)}(\hat{\theta}_m)\}$ and $\{(e_i^{(m)}(\hat{\theta}_m))^2\}$ series, and the test for excess dispersion (ED) based on the $\{e_i^{(m)}(\hat{\theta}_m)\}$ series. $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ denotes the MLE. In addition, the same 3 statistics are also reported using the series of durations obtained by pooling the points associated with $\{e_i^{(1)}(\hat{\theta}_1)\}$ and $\{e_i^{(2)}(\hat{\theta}_2)\}$. In all cases, p-values are calculated using the asymptotic null distributions which hold in the case of known parameters. This approach is in line with most previous work in the econometrics literature. Two alternative feasible procedures that take account of the parameter uncertainty correctly are now presented, 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 ith duration of a univariate PP with $\mathcal{F}_t^N$-intensity $\lambda(t; \theta)$. Then it is possible to show, provided that $\lambda(t)$ can be expressed

[22] 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. The following pragmatic approach to this problem is adopted in Section 5.2. Let C_m be the last observed point of $\tilde{N}_m(t)$ and $C := \min\{C_m, m = 1, ..., M\}$. Then sort $\{\tilde{T}_i^{(m)} : \tilde{T}_i^{(m)} \leq C, m = 1, ..., M\}$ into ascending order and proceed as before.

in a certain form, that[23]

$$F_i(x,\theta) = \Pr[S_i \leq x | S_{i-1}, ..., S_0; \theta] = 1 - \exp\left(- \int_{T_{i-1}}^{T_{i-1}+x} \lambda(s;\theta) ds \right). \quad (21)$$

Clearly, $\{F_i(S_i, \hat{\theta})\} = \{1 - \exp(-e_{i-1}(\hat{\theta}))\}$.[24] 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 the series $\{1 - \exp(-e_{i-1}(\hat{\theta}))\}$. However, this approach is generally not feasible in the case of a multivariate PP model. For example in the bivariate case, letting $S_i^{(m)} = T_i^{(m)} - T_{i-1}^{(m)}$, 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 powerful one.

An alternative procedure that applies to both the univariate and multivariate cases is to minimise the diagnostic test statistic with respect to the parameters. Let $T^{(m)}(\theta_m)$ denote the test statistic evaluated at θ_m (obtained by using the series $\{e_i^{(m)}(\theta_m)\}$). The null of correct specification is rejected if and only if

$$T^{(m)} = \min_{\theta_m \in \Theta_m} T^{(m)}(\theta_m), \quad (22)$$

exceeds the $(1-\alpha)$ quantile of the distribution of $T^{(m)}(\theta_m^*)$ (where θ_m^* is the true parameter value). Since $T^{(m)} \leq T^{(m)}(\theta_m^*)$ everywhere, the probability of falsely rejecting the null is bounded above by α. The validity of the bound does not depend on the PP model satisfying any regularity conditions other than the one in (17) and so the procedure has widespread applicability. Furthermore, the bound is valid when θ_m^* is on the boundary of the parameter space – a case which, as was noted earlier, is of some importance for generalised Hawkes models. Experience has shown that the calculation of such test statistics is computationally feasible for datasets of the size used here and can yield tests with sensible properties.[25]

[23]The condition that $\lambda(t)$ must satisfy and the main part of the proof may be found in Brémaud (1981, E5, p.63).

[24]Note that it is immediate from Corollary 4.2 that $\{F_i(S_i, \theta)\} = \{1 - \exp(-e_{i-1}(\theta))\}$ is *i.i.d.* uniform on $(0,1)$ under P_θ.

[25]It is necessary to choose $T^{(m)}(\theta_m)$ 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.

5 Empirical Application: Trades and Quotes

In order to illustrate the usefulness of generalised Hawkes models in the analysis of financial market event data, this section presents the results of fitting the BV g-HawkesE(2) model to the timing of trades and mid-quote changes for a NYSE stock. The diagnostic testing procedures developed in Section 4 are used to assess the goodness-of-fit of the model. The 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.

5.1 Data

The dataset was extracted from the NYSE Trade and Quote (TAQ) database and records the timing and characteristics of all trades and changes to the mid-quote that occurred on the NYSE for the heavily traded stock General Motors Corporation (GM). The period covered is the 40 trading days from 5 July 2000 to 29 August 2000 inclusive. The 'specialist' for each stock on the NYSE is obliged to report the best quotes (i.e. the highest bid and lowest offer) communicated to the trading crowd and to execute any order at a price that is at least as favourable as his published quote.[26] The mid-quote is defined here as the simple average of the reported bid and ask quotes. A change to the mid-quote occurs when the specialist reports new quotes. A trade occurs when the specialist matches a buy and a sell order (or alternatively, takes one side of the transaction himself). A trade might be expected to affect the waiting time to the next change in the mid-quote because of the information it conveys to the specialist and floor brokers. Similarly, a price change might alter the trade intensity if agents monitor the market closely and submit market orders in order to benefit from advantageous prices. Evidence will be presented below that trades do indeed exert a large impact on the intensity for mid-quote changes, and vice versa.

The adjustments that are made to the data are now described. First, market events (i.e. trades or changes to the mid-quote) occurring outside of normal trading hours (9:30 EST to 16:00 EST) were deleted from the dataset.[27] Having mapped the times of the market events onto $[0, \infty)$ using the data transformation described in Section 3.1, the times of the mid-quote changes were then thinned (i.e. a

[26]The quotes reported by the specialist 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. Further details of the institutional features of the NYSE may be found in Hasbrouck, Sofianos, and Sosebee (1993).

[27]The first reported trade of each calendar day was also removed as this details the price and volume of the opening auction. Trades that were later cancelled were also deleted. 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 was used.

	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 1: 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.

subset of the times was selected) to obtain the 'mid-quote events' that are modelled. A mid-quote event is defined to occur 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.[28] The aim of the analysis here is to examine the two-way interaction of trades and market quotes at a microstructure level of detail. It is therefore desirable not to set the threshold used to thin the quote data too high, in order that changes to the mid-quote over short time horizons are captured. 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. In order to avoid the modelling task becoming overly complicated, the size of the change in the mid-quote between successive events is not modelled directly. 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.[29]

Recall from Definition 2 that, since the pooled process $\{T_i\}$ is always simple, the M-variate PPs considered in this paper assign zero probability to the simultaneous occurrence of two events (of either the same or different types). In the case of the GM dataset, 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 continuous time and the reported times, I adjust the times of those mid-quote events that coincide with a trade by a small, *i.i.d.* uniform amount. The procedure

[28] 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 dataset. 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.

[29] 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.

adopted 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 so as not to impose strong *a priori* assumptions about the original ordering of the two events. A discussion of the treatment of events with identical timestamps is given in the Appendix, together with the results of a sensitivity analysis that examines two further adjustment rules. These results support the robustness of the main empirical findings of this section.

Summary statistics of the intertrade durations and durations between successive mid-quote events for the final dataset are given in Table 1.[30] Note that the average duration between mid-quote events is approximately 6.6 times the average intertrade duration.

5.2 Model estimates and diagnostics

It is well known that intradaily seasonality is an important feature of financial market event data. In common with Russell (1999), it was found that adopting a piecewise linear spline for the deterministic components, $\mu_m(t)$, in (11) worked well in practice. The spline is continuous over the course of the 6.5 hour trading day, with nodes at 9:30,10:00,11:00,...,16:00. Noting that t is measured in hours, $\mu_m(t)$ $(m = 1, 2)$ can be written as:

$$\mu_m(t;\gamma_m) = \begin{cases} 1_{v(t)\in(0,0.5]}[\gamma_{m1} + 2v(t)(\gamma_{m2} - \gamma_{m1})] + \\ \sum_{i=1}^{6} 1_{v(t)\in(i-0.5,i+0.5]}[\gamma_{m,i+1} + (v(t) - i + 0.5)(\gamma_{m,i+2} - \gamma_{m,i+1})] & \text{for } v(t) > 0, \\ \gamma_{m8} & \text{for } v(t) = 0, \end{cases} \tag{23}$$

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 $\gamma_{mi} > 0$ $(i = 1, ..., 8)$. The γ_{mi} $(i = 2, ..., 7)$ are the values of the deterministic component $(i - 1.5)$ hours into each trading day. Note that $\gamma_{m1} \neq \gamma_{m8}$ is allowed.[31]

The results of fitting a restricted BV-g-HawkesE(2) model with the restrictions ($\pi^{(1)}_{mm} = 0$, $\pi^{(1)}_{mq} = 0, \pi^{(2)}_{mq} = 0; m = 1, 2; q = 2$ if $m = 1$ and $q = 1$ if $m = 2$) are shown in Table 2.[32] The restrictions imposed imply that the stochastic components of the intensity exhibit zero spillover effects, except for the $\tilde{\lambda}^{(2)}_{mm}(t)$ $(m = 1, 2)$ components in (11) which capture the $j = 2$ effect of type m events on the mth intensity. The spillover effects are discussed in more detail below. The restricted model fits very

[30]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.

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

[32]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.

$\boldsymbol{\theta}_1$	**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]	$\beta_{21}^{(2)}$
γ_{11}	21.740	[14,34]	74.077	[53,104]	γ_{21}
γ_{12}	2.6181	[0.76,9.0]	45.906	[33,63]	γ_{22}
γ_{13}	0.0000		15.580	[7.9,31]	γ_{23}
γ_{14}	0.0000		21.365	[15,30]	γ_{24}
γ_{15}	0.0000		17.928	[13,25]	γ_{25}
γ_{16}	0.0000		26.408	[21,33]	γ_{26}
γ_{17}	0.7743	[0.14,4.4]	31.456	[25,39]	γ_{27}
γ_{18}	0.4877	[0.01,37]	69.602	[60,80]	γ_{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 2: MLEs and diagnostics for a restricted BV-g-HawkesE(2) model of the timing of trades and mid-quote changes for 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 4.2; 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.

well indeed. Only one of the diagnostic tests in Table 2 rejects at the 5% level.[33] Furthermore, the $(BL, BL(\text{squares}), ED)$ tests based on the durations of the pooled process that is formed *after* the changes of time have been applied (see Section 4.2) were $(23.670, 23.277, -3.7246)$, corresponding to p-values of $(0.0709, 0.0784, 0.0002)$ respectively.

Figure 1 graphs the estimated components of the mid-quote event and trade 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 mth column of the figure ($m = 1, 2$) shows the estimated total

[33] Two sets of diagnostics are reported in Table 2: one for the quote intensity and one for the trade intensity (see Section 4.2 for details).

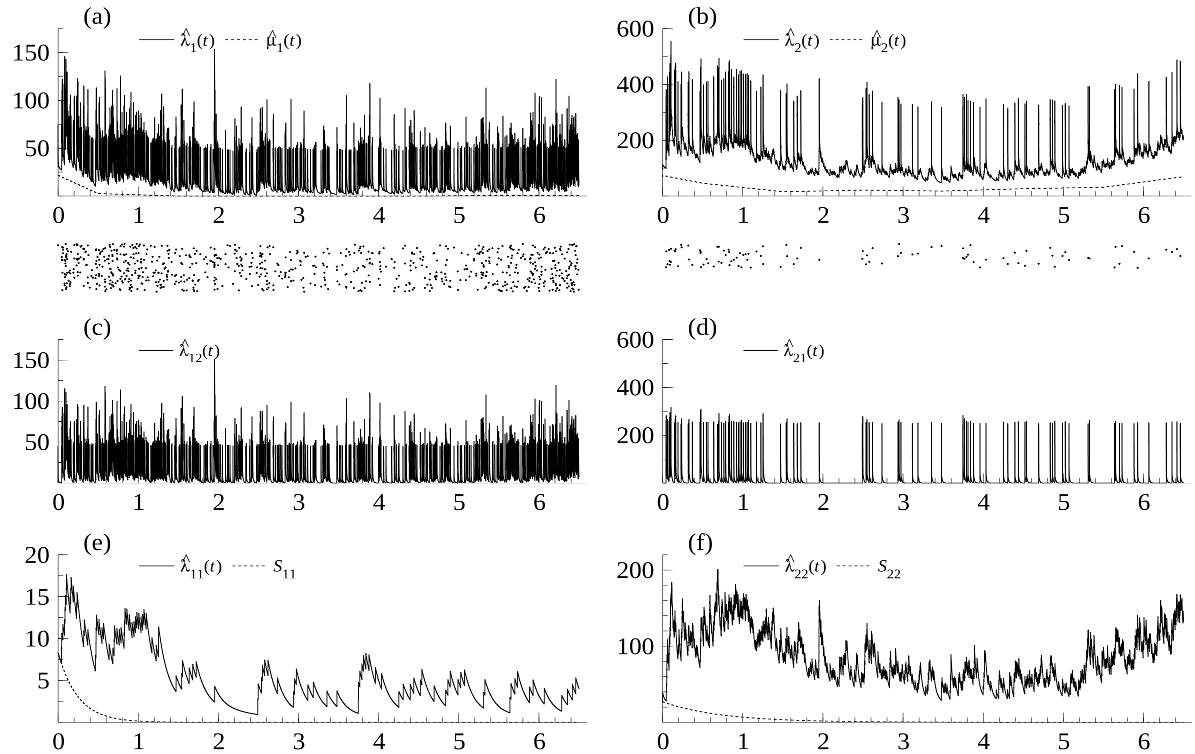

Figure 1: Components of the estimated mid-quote event and trade intensities for General Motors on 2 August 2000. The first column of panels is for the mid-quote event intensity and the second for the trade intensity. (a) the estimated total intensity, $\hat{\lambda}_1(t)$, and estimated deterministic 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.

intensity, $\hat{\lambda}_m(t)$, and estimated deterministic component, $\hat{\mu}_m(t)$; the second shows the occurrence times of the type q 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 (11); 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 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 2. Non-symmetric 95% confidence intervals are shown in square parentheses.[34] The 'U-shape' of the deterministic component of the trade intensity is familiar from previous studies (see, for example, Engle and Russell (1998) and Engle (2000)), whilst the deterministic

[34] The confidence intervals (CIs) were obtained by reparametrising the log-likelihood in terms of $\phi_m = \log(\theta_m)$ $(m = 1, 2)$ and exponentiating the endpoints of the CIs for the elements of ϕ_m. The CIs for ϕ_m were calculated using the inverse of the negative Hessian matrix in the usual way. A pragmatic approach was taken in order to avoid the problems associated with $\gamma_{13}, \gamma_{14}, \gamma_{15}$ and γ_{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.

component of the quote intensity is close to zero after the first hour of the day. Comparing the estimates of $\alpha_{mr}^{(j)}, \beta_{mr}^{(j)}$ ($mr \in \{1,2\} \times \{1,2\}$) for $j = 2$ and $j = 1$, we see that the response function $\alpha_{mr}^{(j)} e^{-\beta_{mr}^{(j)}(s)}$ is initially smaller but 'longer lived' for $j = 2$ than $j = 1$. Of particular economic interest are the estimates of $\alpha_{mq}^{(1)}, \beta_{mq}^{(1)}, \alpha_{mq}^{(2)}$, and $\beta_{mq}^{(2)}$ ($m = 1, 2; q \neq m$). 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 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 1. 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 (11), $\sum_{j=1}^{2} \tilde{\lambda}_{mq}^{(j)}(t)$, 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 1 is one in which the cross effect terms – which capture the effect of type q events on the intensity for type m events ($q \neq m$) – exhibit large, short lived fluctuations that play an extremely important role in determining the dynamics of the process. The economic interpretation of these effects is considered 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. Section 3.3 considered the reasons why the standard LR test cannot be applied in order to test this hypothesis. Nevertheless, it is noted that imposing the restrictions $\alpha_{12}^{(j)} = 0$ ($j = 1, 2$) on the $\lambda_1(t)$ intensity of the (unrestricted) BV-g-HawkesE(2) model yielded 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 2. All 3 diagnostic tests for the mid-quote event intensity continued to accept at the 5% level. Similarly, imposing the restrictions $\alpha_{21}^{(j)} = 0$ ($j = 1, 2$) on $\lambda_2(t)$ resulted in a reduction in the log-likelihood of 450 and in the excess dispersion test then rejecting at the 1% level. The bivariate model thus seems to be a substantial improvement over the two univariate models which ignore the cross effects. Interestingly, the effect of mid-quote events on the intensity for trades should not be ignored. A sensitivity analysis is presented in the Appendix comparing the above results with those obtained using two other adjustment rules for the treatment of trades and mid-quote events with identical timestamps. The analysis strongly suggests that the finding of a positive effect of mid-quote events on the trade intensity is not the result of the particular adjustment rule employed here.

In order to investigate the spillover effects further, two more models were estimated: the BV HawkesE(2) model in (13) and the model with no dependence between trading days given by imposing the restrictions $\pi_{mr}^{(j)} = 0$ $(\forall mr, \forall j)$ on the BV g-HawkesE(2) model. The BV HawkesE(2) model has a log-likelihood of 143,461 (a reduction of 54) and the $ED^{(2)}$ test now has a p-value of 0.018. The model with $\pi_{mr}^{(j)} = 0$ $(\forall mr, \forall j)$ has a log-likelihood of 143,513 (a reduction of 2) and similar diagnostics to those presented in Table 2. In the absence of a formal test of $H_0 : \pi_{mr}^{(j)} = 0$ $(\forall mr, \forall j)$, I have erred on the side of caution and presented the results for the model given in Table 2. However, the hypothesis of no dependence between trading days is not strongly rejected by the data in this case. 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. The failure to reject such a hypothesis is perhaps surprising and is certainly of interest.

5.3 Approximating the instantaneous volatility

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(k) 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 ith mid-quote event as usual by $T_i^{(1)}$ and the associated mid-quote change by Δ_i. The (right continuous) price process can thus be written as $P(t) = P(0) + \sum_{i:T_i^{(1)} \leq t} \Delta_i$. Then define the instantaneous conditional volatility by

$$\sigma^2(t) = \lim_{h \downarrow 0} \mathrm{E}\left[\frac{1}{h}\left(\frac{P(t+h) - P(t)}{P(t)}\right)^2 | \mathcal{F}_t\right], \tag{24}$$

where $\mathcal{F}_t = \sigma(P(t)) \vee \mathcal{F}_t^N$ and N is the bivariate PP of trades and mid-quote events. As Engle and Russell (1998) have noted, it is possible to express $\sigma^2(t)$ in terms of the mid-quote intensity. Specifically,

$$\sigma^2(t) = c^2 \lambda_1(t+)/P(t)^2, \tag{25}$$

where Aalen (1978, Lemma 3.3(ii)) has been used and it has been 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).

An approximate estimate of the instantaneous volatility can be formed using the estimates for the

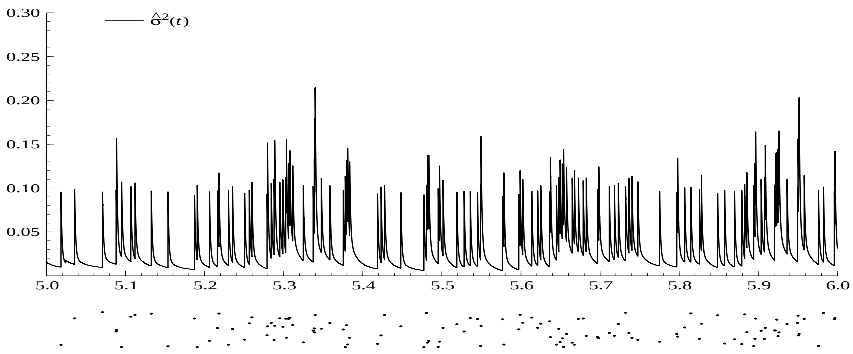

Figure 2: Estimated instantaneous volatility, $\hat{\sigma}^2(t)$, for General Motors on 2 August 2000 between 14:30 and 15:30 EST. $\hat{\sigma}^2(t)$ is based on equation (25). Also shown are the (vertically jittered) occurrence times of the trades. Volatility is measured per trading year (where one year is 252×6.5 trading hours). The horizontal axis is time measured in hours since the start of the trading day.

restricted BV-g-HawkesE(2) model by making the following substitutions in (25): 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 ith mid-quote event for the dataset and Δ_i is the actual change to the mid-quote since the last mid-quote event).[35] This estimate is graphed in Figure 2 for the trading hour between 14:30 and 15:30 EST on 2 August 2000 (the same day as that used in Figure 1). 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 were 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 direction of the price changes between mid-quote events.

5.4 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

[35] There are two 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, (25) 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.

contrast, the modelling framework presented here allows the two-way interaction of trades and mid-quote changes to be analysed. The empirical results of Sections 5.2 and 5.3 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 prices.[36] 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 procedure. 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. 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 the 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 O'Hara (1995, Ch. 2)), the altered quote will tend to be 'hit' soon

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

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 stock price over time and hence reduced incentives for the informed to trade. Such an effect is not possible in the BV g-HawkesE(k) model since the constraints $\alpha_{21}^{(j)} \geq 0 \; \forall j$ are imposed. As is discussed briefly in Section 6.2, it is possible to allow for such negative effects in the generalised Hawkes modelling framework and such an extension would be of some interest given the findings here.

For the estimated BV g-HawkesE(2) 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 2). 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 are few trades. A number of other empirical studies have reported similar findings.[37] 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. This is in line with the finding here that the occurrence of a mid-quote change results in a large increase in the trade intensity. In contrast, Grammig and Wellner (2002) find that higher lagged volatility significantly reduces trade intensity.[38] The further investigation of the effect of prices on the trade arrival process is an interesting challenge for both theoretical and empirical microstructure research.

6 Comparison with previous research and possible extensions

There appear to be only two other papers that consider empirical models of multivariate financial market event data set in continuous time. These are discussed in turn below, before indicating some possible

[37] 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.

[38] 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.

extensions to this work and directions for future research.

6.1 Other multivariate models

Engle and Lunde (2003) (hereafter EL) specify models for a bivariate sequence of durations $\{S_i^{(1)}, S_i^{(2)}\}_i$, where $S_i^{(1)}$ is the ith intertrade duration and $S_i^{(2)}$ is the time from the start of the ith intertrade duration until the next trade or mid-quote change (whichever occurs first). The trade arrival process is treated as the 'driving process', thus avoiding the difficulty that would otherwise arise with the origins of $S_i^{(1)}$ and $S_i^{(2)}$ possibly being far apart in real time. The EL model is rather specialised, being designed to analyse the speed with which the information content of trades is incorporated into prices. This sharp focus does, however, limit its scope and wider applicability. 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.[39] It is worth noting that a full M-variate PP model may be specified via $\mathrm{P}[T_{i+1} - T_i \leq x, Z_{i+1} = m | \mathcal{F}_{T_i}^N]$ where $\mathcal{F}_{T_i}^N = \sigma(T_1, Z_1, T_2, Z_2, .., T_i, Z_i)$ – that is, via the joint conditional probability of the next duration and mark of the *pooled process*.[40]

The approach adopted in this paper is closest to that of Russell (1999), whose Autoregressive Conditional Intensity (ACI) model is also, as the name suggests, specified via the stochastic intensity. For expositional purposes, there follows a definition of the simplest bivariate ACI(1,1) model (without time varying intensity between events). The model is an M-variate PP model (with $M = 2$). The $\mathcal{F}_t^N$-intensity of the process is given by

$$\lambda_m(t) = \exp(\phi_{m,i}), \quad T_i < t \leq T_{i+1} \tag{26}$$

for $i = 0, 1, ...$, where $\phi_i = (\phi_{1,i}, \phi_{2,i})'$ has the autoregressive specification

$$\phi_i = \begin{cases} \eta_1 \varepsilon_i + B\phi_{i-1} \text{ if } Z_i = 1 \\ \eta_2 \varepsilon_i + B\phi_{i-1} \text{ if } Z_i = 2 \end{cases}, \tag{27}$$

with η_1, η_2 and B parameters of the appropriate dimension. Note that i indexes the events of the pooled process $\{T_i\}$. Let $T_i^* := \sup\{T_j : j < i \text{ and } Z_j = Z_i\}$ provided that this set is not empty and let $T_i^* := 0$

[39] 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.

[40] Kamionka (2001) and Spierdijk, Nijman, and van Soest (2002) both essentially adopt this approach to specify bivariate PP models.

otherwise.[41] Then, the r.v. $\varepsilon_i(\omega)$ is given by

$$\varepsilon_i(\omega) = 1 - \tilde{\varepsilon}_i(\omega) = 1 - \int_{T_i^*(\omega)}^{T_i(\omega)} \lambda_{Z_i(\omega)}(\omega, s)ds. \tag{28}$$

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 4.2 since λ is the $\mathcal{F}_t^N$-intensity). The series $\{\phi_i\}$ is thus an AR(1) process. This way of specifying the model allows the methods of time series analysis to be brought into play. However, I have been unable to prove that the ε_i in (28) are independent (or even uncorrelated) in the bivariate case. It does *not* follow from Theorem 4.1 and Corollary 4.2 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 4.2, it can be seen that although a particular realisation, $\{\tilde{\varepsilon}_i(\omega)\}$, 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 ω). It may well be that further work will resolve this problem.[42] An alternative model is given by leaving the specification unchanged except that ε_i is now given by

$$\varepsilon_i = 1 - \tilde{\varepsilon}_i = 1 - \int_{T_{i-1}}^{T_i} \sum_{m=1}^{2} \lambda_m(s)ds. \tag{29}$$

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 4.1 and Corollary 4.2 that $\{\tilde{\varepsilon}_i\}$ is *i.i.d.* $Exp(1)$ in this case.

To sum up, the ACI-type model structure offers alternative model specifications to the generalised Hawkes models developed here. An advantage of the ACI framework is the rich variety of time series specifications that could be adopted for ϕ_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, both families of models should be judged by their ability to provide well specified empirical models whose parameters may be readily interpreted in economic terms.

6.2 Possible extensions

An important strength of the approach to model specification, estimation and diagnostic testing that has been described in this paper is that it is straightforward to use the same methods when the conditioning is

[41] T_i^* denotes the time of the most recent event of the pooled process that occurred prior to the ith event and was of the same type as that event.

[42] The author 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.* $Exp(1)$.

on a wider filtration, $\{\mathcal{F}_t\}$, with $\mathcal{F}_t \supset \mathcal{F}_t^N \ \forall t$.[43] Important additional information that might be included in $\mathcal{F}_t$ in this context includes the characteristics or 'marks' of the events other than the event type (for example, the size and direction of trades), and data concerning news announcements and 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. For example, consider the specification of the term $\tilde{\lambda}_{12}^{(j)}(t)$ in equations (11) and (12), and suppose that type 1 (resp. type 2) events are mid-quote changes (resp. trades). In (12), the jump in $\tilde{\lambda}_{12}^{(j)}(t)$ in response to a trade is always equal to $\alpha_{12}^{(j)}$. The suggested extension is to set the jump that occurs in response to the ith trade equal to Y_i, where Y_i is a parametrised function of the marks of that trade (denoted here by $Z_i^{(2)}$). The BV-g-HawkesE(k) type specification for $\tilde{\lambda}_{12}^{(j)}(t)$ thus becomes

$$\tilde{\lambda}_{12}^{(j)}(t) = \pi_{12}^{(j)} \tilde{\lambda}_{12}^{(j)}(\tau_{d-1}) \mathrm{e}^{-\rho_{12}^{(j)}(t-\tau_{d-1})} + \sum_{i:T_i^{(2)} \in [\tau_{d-1}, t)} Y_i \mathrm{e}^{-\beta_{12}^{(j)}(t-T_i^{(2)})}, \tag{30}$$

for $\tau_{d-1} < t \le \tau_d$ $(d = 1, 2, ...)$, where $Y_i = g(Z_i^{(2)})$ for some function g. Such an extension should prove useful in future market microstructure research.[44]

This paper points to a number of directions 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 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.[45] Finally, an investigation of the power properties of diagnostic tests computed by minimising over the parameter space, as described in Section 4.2, would be worthwhile.

[43]The econometrician then specifies the $\mathcal{F}_t$-intensity of $N_m(t)$ for $m = 1, ..., M$. Theorem 4.1 requires that $\mathcal{F}_t \supseteq \mathcal{F}_t^N \ \forall t$. The diagnostic testing procedures therefore continue to be valid when $\mathcal{F}_t \supset \mathcal{F}_t^N \forall t$.

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

[45]Ogata (1981) extends Lewis' thinning simulation algorithm to multivariate PPs which are absolutely continuous 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.

7 Conclusion

This paper has developed 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 specification, 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 stochastic 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 provides a useful framework for approaching the issue of dependence between trading days. Furthermore, the structure allows the intensity at the start of the trading day to depend on additional information such as overnight news announcements. Analytic likelihoods are available for the generalised Hawkes models and MLEs can readily be computed.

It has been shown how to use random changes of time to construct diagnostic tests for 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 natural in the context of financial market event data. Novel diagnostic tests are also suggested that utilise the superposition of the (independent) Poisson processes obtained by the random changes of time, thus permitting a more rigorous check of the model specification to be conducted overall. 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.

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(k) model allows the study of the two-way interaction between trades and price changes. It is found that not only do trades 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 including the ACI model of Russell (1999) is provided, and an extension of the generalised Hawkes models which conditions on the additional characteristics or marks of the events is proposed.

8 Appendix

8.1 Adjustment rules for simultaneous events

The M-variate PPs in this paper assign zero probability to the simultaneous occurrence of two market events (of either the same or different types). This section describes the adjustments that were made to the General Motors (GM) dataset analysed in Section 5 in order to deal with such simultaneous events. Events of the same type (i.e. two trades or two quotes) that have the same timestamp were treated as a single event with that timestamp.[46] The occurrence of such events was rare.[47] The treatment of the simultaneous occurrence of trades and mid-quote events is a more substantive issue. Approximately 11 per cent of the mid-quote events have exactly the same timestamp as a trade. In Section 5, I have adjusted 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 $(0, 1)$. 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. The actual occurrence time refers to 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 close succession 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

[46]Note that the timestamps of all events in the original data are an integer number of seconds.
[47]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.

being reported with identical timestamps as a result of lengthier reporting delays. I 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.[48]

Since it would be expected that the estimates of $\alpha_{mq}^{(1)}, \beta_{mq}^{(1)}, \alpha_{mq}^{(2)}, \beta_{mq}^{(2)}$ $(m = 1, 2; q \neq m)$ are sensitive to the particular treatment of the identical timestamps that is adopted, a sensitivity analysis was conducted comparing the 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 2. It is the positive effect of mid-quote events on the trade intensity that is perhaps the more surprising of the empirical findings in Section 5.2. 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. Thus, the finding of a positive effect of mid-quote events on the trade intensity should not be interpreted as the result of having employed a data adjustment rule that was unduly biased in favour of finding such an effect.

[48] 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 close together in real time. 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.

8.2 Computation of MLEs

In order to simplify the presentation, I first describe the procedure used to compute MLEs for the univariate g-HawkesE(k) models and then discuss how to extend this procedure to the bivariate case. Evaluating the log-likelihood for the univariate g-HawkesE(k) model yields

$$
\begin{aligned}
l(\theta; T_1, ..., T_{N(T)}) \quad = \quad & -\int_0^T \mu(s\ ;\gamma)ds - \sum_{d=1}^{T/l}\sum_{j=1}^{k}\Big\{\pi_j/\rho_j(1-\mathrm{e}^{-l\rho_j})\tilde{\lambda}_{j,\theta}(\tau_{d-1}) + \\
& \sum_{\tau_{d-1}\le T_i<\tau_d} \alpha_j/\beta_j(1-\mathrm{e}^{-\beta_j(\tau_d-T_i)})\Big\} + \sum_{d=1}^{T/l}\sum_{T_i\in A_d}\log\{\mu(T_i\ ;\gamma) + \\
& \sum_{j=1}^{k}\Big[\pi_j\tilde{\lambda}_{j,\theta}(\tau_{d-1})\mathrm{e}^{-\rho_j(T_i-\tau_{d-1})} + \alpha_j A_{d,j}(i)\Big]\} + T,
\end{aligned} \tag{31}
$$

where $A_{d,j}(i) = \sum_{z:\tau_{d-1}\le T_z<T_i} \mathrm{e}^{-\beta_j(T_i-T_z)}$ for $d = 1, 2, ..., T/l$ (and the convention that an empty sum is equal to zero is adopted). When computing (31) in practice, the recursion in (10) is used to calculate the $\tilde{\lambda}_{j,\theta}(\tau_{d-1})$'s. Use is also made of the following recursion in i in order to improve computational efficiency

$$
A_{d,j}(i+1) = (1 + A_{d,j}(i))\mathrm{e}^{-\beta_j(T_{i+1}-T_i)}. \tag{32}
$$

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 $\delta_j = \beta_j/\beta_{j-1}$ $(j = 2, ..., k)$. The constraints $\alpha_j > 0$, $\pi_j > 0, \rho_j > 0$ $(j = 1, ..., k)$, $\gamma_1 > 0, ..., \gamma_8 > 0$, and $\beta_1 > \beta_2 > ... > \beta_k > 0$ are thus imposed.

The evaluation of the log-likelihood for the bivariate g-HawkesE(k) model given by (14) and (15) is a straightforward extension of (31) above. When performing numerical optimisation, the log-likelihood was reparametrised in the same way as for the univariate model (with the obvious extension to the bivariate case). A slight complication arises in the need to calculate terms of the form

$$
A_{mq}^{(d,j)}(i) = \sum_{z:\tau_{d-1}\le T_z^{(q)}<T_i^{(m)}} e^{-\beta_{mq}^{(j)}(T_i^{(m)}-T_z^{(q)})}, \tag{33}
$$

for all $T_i^{(m)} \in (\tau_{d-1}, \tau_d]$ (where $q = 2$ if $m = 1$ and $q = 1$ if $m = 2$). The following recursion in i is useful in this regard:

$$
A_{mq}^{(d,j)}(i+1) = A_{mq}^{(d,j)}(i)\mathrm{e}^{-\beta_{mq}^{(j)}(T_{i+1}^{(m)}-T_i^{(m)})} + \sum_{z:T_i^{(m)}\le T_z^{(q)}<T_{i+1}^{(m)}} e^{-\beta_{mq}^{(j)}(T_{i+1}^{(m)}-T_z^{(q)})}. \tag{34}
$$

When estimating the univariate g-HawkesE(2) model, initial parameter values for the MaxBFGS algorithm were obtained by estimating the following sequence of models:[49]

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, I maximise $l_m(\theta_m) - T$ (where $l_m(\theta_m)$ is given by (15)) for $m = 1$ and $m = 2$ separately. This reduces the dimensionality of the parameter space of the optimisation problem and is valid since $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_1$. 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. The starting values used were the estimates obtained from the univariate HawkesE(2) model for type m events in step 3) above:

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.

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

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

[49]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.

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.

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. (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.

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.