Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models

Clive G. Bowsher*

August 23, 2006

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 conditional intensity. Generalised Hawkes models are introduced that incorporate inhibitory events and dependence between trading days. Novel omnibus specification tests based on a multivariate random time change theorem are proposed. A bivariate point process model of the timing of trades and mid-quote changes is then presented for a New York Stock Exchange stock and related to the market microstructure literature. The two-way interaction of trades and quote changes in continuous time is found to be important empirically.

Keywords: Point process, conditional intensity, Hawkes process, specification test, random time change, transactions data, market microstructure.

JEL classification: C32, C51, C52, G10.

1 Introduction

This paper develops a continuous time econometric modelling framework for analysis of the dynamic microstructure of financial markets – that is, their dynamic evolution viewed at a very fine level of detail. These dynamics can be described in terms of the stochastic occurrence times and characteristics of well-defined *market events* such as trades and changes to the quoted prices. The work is motivated by the growing theoretical market microstructure literature and the advent of datasets providing complete records for some or, in the case of certain electronic markets, all types of market event. Such datasets potentially provide an enormous amount of information about the intraday behaviour of financial markets and allow testing of the hypotheses of the theoretical literature. However, progress in modelling the data in continuous time has hitherto been hindered by the difficulties presented by the multivariate case. Furthermore, many interesting economic questions concerning financial market microstructure can only be addressed using such multivariate models. The approach adopted here overcomes these difficulties by focusing on the *conditional intensities* of the market event arrival processes, and provides a general

*Nuffield College, University of Oxford, New Road, Oxford, OX1 1NF, UK, *tel:* +44 1865 278969, *fax:* +44 1865 278621, *e-mail:* clive.bowsher@nuffield.ox.ac.uk. An earlier long version of this paper appeared in working paper form in Bowsher (2002). Particular thanks are due to Charles Bos, David Cox, Patrick Hewlett, Jeremy Large, Bent Nielsen, Jeff Russell, Mark Salmon, Neil Shephard, Matthias Winkel and 3 anonymous referees. Any remaining errors are my own. All computational work was performed using the Ox language (see Doornik 2001). Financial support from the ESRC under awards R42200034061 and R00023839 is gratefully acknowledged.

framework for model specification and inference that it is hoped will greatly facilitate the econometric analysis of these vast and important datasets in the future.

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 models that permits the analysis of *multivariate* financial market event (or 'transactions') data – that is, data that records the timing and characteristics of several different *types* of market event.[1] The models are general enough to incorporate 'inhibitory' events that result in a decrease in a conditional intensity and to allow dependence between trading days. Second, the use of a multivariate random change of time to construct specification tests for parametric point process models is established. Novel omnibus tests for the multivariate case are proposed together with tests of the specification of each component, scalar conditional intensity. A sufficient condition for the validity of the testing procedures is derived which is shown to be natural in the context of financial markets. Finally, applying the econometric methods developed in the paper to data for 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.

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 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.

In the econometric analysis, the market event data are 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 intervals or durations between events (see in particular, Engle and Russell 1997,1998),

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

multivariate extensions of this work have been slow to emerge in the econometrics literature.[2] Engle and Lunde (2003) specify a model 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. It is worth noting that the model does not (and does not claim to) imply a conditional intensity in continuous time for mid-quote change events.[3] The specification of a full multivariate PP model may be achieved within a duration-based approach by specifying the joint conditional probability of $(T_{i+1}-T_i)$ and Z_{i+1}, that is the next duration and event type respectively of the PP, given its observed history up to (and including) the ith event.[4] However, progress in this direction has so far been limited – indeed, the creation of new empirical models from this starting point seems unduly difficult.

I adopt a different approach in which the model is specified via the conditional intensity for each type of market event. This provides a powerful and more natural modelling framework for multivariate market event data. The common information set upon which each intensity is conditioned is updated continuously as new information arrives, thus allowing other types of event to have an immediate impact on the intensity as they occur in continuous time.[5] When the intensities are conditioned on the natural filtration of the multivariate PP, the approach is mathematically equivalent to the duration-based one just discussed but is intuitively more appealing. The intensity-based approach also has the advantage that it is straightforward to condition additionally on the history of some other continuous time process. Intensity-based modelling in the econometrics literature has hitherto utilised the Autoregressive Conditional Intensity (ACI) model of Russell (1999) and its extensions. Hall and Hautsch (2004) estimate a bivariate ACI model for the arrival of buy and sell trades on a limit order book market. Bauwens and Hautsch (2003) propose an extension of the ACI model which adds a latent, Gaussian autoregressive component to the log intensity. The core component of all of these ACI models is specified in 'event time', with the consequence that understanding the properties of the continuous time conditional intensity process, or equivalently the

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

[3]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.

[4]Note that the duration in mind here is that of the so-called 'pooled process', $\{T_i\}$. Spierdijk, Nijman, and van Soest (2004) adopt this approach to a multivariate study of the timing of trades in several stocks of the same sector.

[5]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.

distribution of the multivariate PP, is difficult.

The structure of the present paper is as follows. Section 2 introduces the new generalised Hawkes (g-Hawkes) models for financial market event data and establishes sufficient conditions for identification of linear Hawkes models. Section 3 describes an intensity-based approach to likelihood inference for PPs and considers the maximum likelihood estimation (MLE) of g-Hawkes models. In novel work, Section 4 proposes a range of specification tests based on a multivariate random time change theorem for PPs. Section 5 then makes use of the models and specification testing procedures developed earlier in order to analyse the two-way interaction between the timing of trades and mid-quote changes for a New York Stock Exchange (NYSE) stock. The empirical findings are related to both the theoretical and empirical market microstructure literature. Section 6 develops a computationally efficient algorithm for the simulation of g-HawkesE(k) processes and reports the results of a Monte Carlo study of the properties of the MLEs and specification tests for the bivariate g-HawkesE(k) model used in the empirical analysis of Section 5. Section 7 then concludes.

Diffusions and point processes are two major classes of continuous time process, but the latter are still far less familiar in econometrics. Appendix B therefore briefly provides some technical background on point processes, counting processes and the martingale-based definition of a conditional intensity. The names of mathematical objects in the theory of point processes are italicised the first time they occur in Section 2 and, if not rigorously defined in the main text, are given proper definition in Appendix B.

2 Generalised Hawkes Models

This section introduces the generalised Hawkes models – a class of models for multivariate market event data which are specified via the vector conditional intensity. Recall that the type of dataset in mind here consists of a record of the occurrence time and type of each market event, the events being sequenced according to the order in which they occurred. For example, Section 5 uses a bivariate model to analyse the relationship between the timing of two different types of market event, namely trades and price (mid-quote) changes.

As noted in the introduction, the events will be described as the realisation of an *M-variate point process* $\{T_i, Z_i\}_{i\in\{1,2,...\}}$, where T_i denotes the occurrence time of the ith event and $Z_i \in \{1, 2, ..., M\}$

indicates the ith event's type.[6] The *M-vector counting process* associated with $\{T_i, Z_i\}$ will be written as $N(t) := (N_m(t))_{m=1}^M$, with each element $N_m(t)$ counting the number of *type m events* that have occurred up to and including time t.[7] Note that the different types of event are indexed here by m and that $T_i^{(m)}$ is used to denote the occurrence time of the ith type m event. The *internal history* (or *natural filtration*) of the M-variate PP $N(t)$ is denoted by $\{\mathcal{F}_t^N\}_{t\geq 0}$, the information set $\mathcal{F}_t^N$ corresponding to complete observation of $N(t)$ up to (and including) time t.

The distinguishing feature of the approach here is that the M-variate PP model is specified via the vector conditional intensity process, $\lambda(t) = (\lambda_m(t))_{m=1}^M$. This perspective is very natural and productive when developing new empirical models – one considers how the intensity for type m events, $\lambda_m(t)$, might change in continuous time as new information arrives. Furthermore, it will be shown later that such an intensity-based approach readily lends itself to both likelihood inference and specification testing using a random change of time. Each intensity, $\lambda_m(t)$, can be interpreted as the conditionally expected number of type m events per unit time as the time interval tends to zero. When the conditioning filtration is $\{\mathcal{F}_t\}$ and the data generating process (DGP) is P, we say that $\lambda(t)$ *is the* $(\mathsf{P}, \mathcal{F}_t)$*-intensity of* $N(t)$. Note that the term $(\mathsf{P}, \mathcal{F}_t)$-intensity is always used here in the sense of the martingale-based definition given in Definition 3 of Appendix B.2. The reader interested in the technical details is referred to that section of the paper. It is important always to be clear about which conditioning filtration is in mind – this section and much of the paper is concerned with $\mathcal{F}_t^N$-intensities that are defined with respect to the internal history, $\{\mathcal{F}_t^N\}$.

The remainder of this section describes various bivariate models in the generalised Hawkes class. Section 2.2 introduces the linear g-HawkesE(k) models which are applied later in the empirical analysis of Section 5, and establishes sufficient conditions for identification of linear Hawkes models. Section 2.3 then extends these models to give new non-linear specifications allowing for inhibitory events, that is events that result in a decrease in a conditional intensity.[8] The concepts introduced can readily be extended to the general multivariate case ($M \geq 2$). The bivariate models have two important features. The $\mathcal{F}_t^N$-intensity for type 1 events (e.g. changes to the mid-quote), $\lambda_1(t)$, depends on the history of type

[6] It is assumed throughout that no two occurrence times are the same (i.e. that the PP $\{T_i\}$ is *simple*) – whilst this feature can sometimes be restrictive in empirical applications, it should be noted that almost all of the existing probability and statistics literature in the field of point processes makes this assumption.

[7] 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, $N(t)$'.

[8] A more flexible nomenclature than g-HawkesE(k) for the *linear* models would be L-gHawkesE(k), but the distinction is not needed in the sequel since the non-linear models feature only in Section 2.3.

2 events (e.g. trades) 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. Finally, Section 2.4 discusses how to condition also on covariates or 'marks' associated with the events. In order to clarify the presentation that follows, it is useful first to discuss the data transformation that is used throughout the paper.

2.1 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 separated by non-trading periods.[9] The approach taken here is to concatenate the data pertaining to each of the days of length l in order to remove the non-trading 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 in the dataset ($d = 1, 2, ...$), then that event has occurrence time $x + l(d - 1)$ in the final dataset. The real half-line is thus partitioned as follows

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

with $\tau_d = l \cdot d$ ($d = 0, 1, 2, ...$) representing the end of the dth trading day. Treating the data as a realisation of a single PP on $(0, T]$ allows the use of existing theorems in the PP literature. However, it is important then to model carefully how the intensity depends on the events of previous calendar days.[10]

2.2 Generalised HawkesE(k) models

In order to simplify the exposition, the univariate, linear g-HawkesE(k) process will be described before then extending the discussion to the bivariate case. Its intensity 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.

Definition 1 ***Univariate g-HawkesE(k) process.*** *The process is defined by the (scalar) $\mathcal{F}_t^N$-conditional intensity*

$$\lambda(t) = \mu(t) + \sum_{j=1}^{k} \tilde{\lambda}_j(t) \qquad \forall t \geq 0, \tag{1}$$

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

[10]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 days), but this can be achieved within the 'single point process' framework adopted here.

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

$$\tilde{\lambda}_j(t) = \pi_j \tilde{\lambda}_j(\tau_{d-1}) e^{-\rho_j(t-\tau_{d-1})} + \int\limits_{[\tau_{d-1},t)} \alpha_j e^{-\beta_j(t-u)} dN(u), \tag{2}$$

for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ...)$, where $\alpha_j \geq 0, \beta_j > 0, \pi_j \geq 0$, and $\rho_j > 0$.[11]

The conditional intensity of the g-HawkesE(k) process is thus the sum of a deterministic component, $\mu(t)$, and k stochastic components, $(\tilde{\lambda}_j(t))_{j=1}^k$. Equation (2) 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 (2) as the (jth) intensity 'spillover effect' between trading days. The purpose of including such terms in the model specification is discussed below.

Evaluating the second term in (2) yields $\sum_{i:\tau_{d-1} \leq T_i < t} \left[\alpha_j e^{-\beta_j(t-T_i)}\right]$, where the 'response function' $\left[\alpha_j e^{-\beta_j(t-T_i)}\right]$ gives the (jth) impact at time t of the event with occurrence time T_i (as a function of t). The term has sample paths that are left-continuous, jumping up by an amount α_j in response to the occurrence of an event and then decaying until the occurrence of the next event. The events are thus self-exciting in the sense that their occurrence increases the intensity for future events, resulting in their 'clustering' which is a well known feature of market event data. When $k > 1$, the superposition of the stochastic components $(\tilde{\lambda}_j(t))_{j=1}^k$ in (1) thus introduces additional flexibility in the shape of the response of the conditional intensity to past events.[12] In practice, the superposition is important empirically since it allows events to exert 'separable' short and long lived impacts on the conditional intensity (see Section 5.2).

The univariate g-HawkesE(k) process can readily be extended to the bivariate (BV) case by including k 'cross effect' terms that capture the effect of type 2 events on the intensity for type 1 events, and vice versa.

Definition 2 ***Bivariate (BV) g-HawkesE(k) process.*** *The process is defined by the vector $\mathcal{F}_t^N$-conditional intensity $(\lambda_1(t), \lambda_2(t))'$, where*

$$\lambda_m(t) = \mu_m(t) + \sum_{j=1}^{k} \tilde{\lambda}_{mm}^{(j)}(t) + \sum_{j=1}^{k} \tilde{\lambda}_{mq}^{(j)}(t), \tag{3}$$

[11] Note the integral with respect to the counting process $N(u)$ in (2). Such integrals are explained in Appendix B.1.

[12] This follows since the sum of k Exponential functions, as in the 'E(k)' models, is not itself exponential. Other specifications of the response functions, including hyperbolic functions, are clearly possible but are not considered here.

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

$$\tilde{\lambda}^{(j)}_{mr}(t) = \pi^{(j)}_{mr}\tilde{\lambda}^{(j)}_{mr}(\tau_{d-1})e^{-\rho^{(j)}_{mr}(t-\tau_{d-1})} + \int\limits_{[\tau_{d-1},t)} \alpha^{(j)}_{mr}e^{-\beta^{(j)}_{mr}(t-u)}dN_r(u), \tag{4}$$

for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ...)$, $\tilde{\lambda}^{(j)}_{mr}(0) = 0$, where $mr \in \{1, 2\} \times \{1, 2\}$. The parameter constraints $\alpha^{(j)}_{mr} \geq 0$, $\beta^{(j)}_{mr} > 0$, $\pi^{(j)}_{mr} \geq 0$, $\rho^{(j)}_{mr} > 0$ $(\forall mr$ and $\forall j)$ apply.

Note the presence of the cross effect terms $\tilde{\lambda}^{(j)}_{mq}(t)$ in (3), which allow the occurrence of type q events to influence the intensity for type m events.[13] 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 (2) and (4). The BV g-HawkesE(k) process nests two important cases. First, when $\pi^{(j)}_{mr} = 0$ ($\forall mr$ and $\forall j$) there is no dependence between trading days since there are then no intensity spillover effects for either $\lambda_1(t)$ or $\lambda_2(t)$. Second, when the restrictions given by Proposition 2.1 below are in force, the BV g-HawkesE(k) process is identical to the bivariate 'mutually exciting' process of Hawkes (1971). This process – referred to here as the 'BV linear Hawkes (1971)' process – can be written as $\lambda_m(t) = \mu_m(t) + \breve{\lambda}_m(t)$ where

$$\breve{\lambda}_m(t) := \sum_{r=1}^{2}\sum_{j=1}^{k} \int\limits_{(0,t)} \alpha^{(j)}_{mr}e^{-\beta^{(j)}_{mr}(t-u)}dN_r(u), \tag{5}$$

for $m = 1, 2$, $\mu_m(t) > 0$, and $\alpha^{(j)}_{mr} \geq 0, \beta^{(j)}_{mr} > 0$. Note that the integral in (4) is the same as that in (5) except that the range of the former is restricted to $[\tau_{d-1}, t)$.

Proposition 2.1 *The BV g-HawkesE(k) process with the parameter restrictions ($\pi^{(j)}_{mr} = 1$, $\rho^{(j)}_{mr} = \beta^{(j)}_{mr}$ $\forall mr$, $\forall j$) imposed and the BV linear Hawkes (1971) process are identical.*

Proof. *A consequence of the fact that for all $mr \in \{1, 2\} \times \{1, 2\}$, the process $\{\tilde{\lambda}^{(j)}_{mr}(t)\}_{t>0}$ in (4) can be written as*

$$\tilde{\lambda}^{(j)}_{mr}(t) = \sum_{s=1}^{d-1}(\pi^{(j)}_{mr})^s C^{(j)}_{mr}(d-s)e^{-\rho^{(j)}_{mr}(t-l(d-s))} + \int\limits_{[\tau_{d-1},t)} \alpha^{(j)}_{mr}e^{-\beta^{(j)}_{mr}(t-u)}dN_r(u), \tag{6}$$

for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ...)$ where

$$C^{(j)}_{mr}(d) = \int\limits_{[\tau_{d-1},\tau_d)} \alpha^{(j)}_{mr}e^{-\beta^{(j)}_{mr}(\tau_d-u)}dN_r(u), \tag{7}$$

is the contribution of the type r events occurring on day d. The result follows by imposing the parameter restrictions on (6) and some algebraic manipulation. ■

[13] By definition, $q \neq m$. Note that both $\tilde{\lambda}^{(j)}_{mm}(t)$ and $\tilde{\lambda}^{(j)}_{mq}(t)$ are defined by equation (4).

This proposition is interesting since it shows that applying a linear Hawkes (1971) model to concatenated trading periods implies spillover effects between periods based on the end-of-period levels $\tilde{\lambda}^{(j)}_{mr}(\tau_{d-1})$ and equal to $[1 \cdot \tilde{\lambda}^{(j)}_{mr}(\tau_{d-1})\mathrm{e}^{-\beta^{(j)}_{mr}(t-\tau_{d-1})}]$. This has not been recognised before. The g-HawkesE(k) model does not necessarily impose the restrictions of Proposition 2.1, thereby providing a much more flexible specification of the spillover effects than the original Hawkes (1971) model.

2.2.1 Recursive model specification

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 g-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 non-trading periods), without taking the special nature of the times $(\tau_1, \tau_2, ...)$ that correspond to the ends of the trading days into account.[14] By adopting a flexible specification, the g-HawkesE(k) models nest both of these approaches whilst also allowing for more general spillover effects. Empirically, such generality is found to be an important aspect of the model specification - see the discussion in Section 5. The second motivation for the recursive specification is that it 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 (3), 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 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 g-Hawkes models, for example, one in which the effect depends on the entire path of the component of the intensity during the previous day via the term $\int_{\tau_{d-2}}^{\tau_{d-1}} W(\tau_{d-1} - s)\tilde{\lambda}^{(j)}_{mr}(s)ds$, where $W(.) \geq 0$ is a non-negative 'weighting' function.

[14] In the latter case the same model (e.g. the linear Hawkes (1971) model) is fitted to several days of data as is used for a single day or in the case of continuous trading.

A further extension would be to allow the spillover effect to depend on the length of the non-trading periods, which are variable due to weekends and holidays.

2.2.2 Identification of linear Hawkes models

Conditions for the statistical identification of linear HawkesE(k) models, including the Hawkes (1971) models, do not appear to have been considered previously in either the statistics or econometrics literature. The pertinent issue is whether these models are identified for any finite k. Since the proof is already quite complicated for a BV linear Hawkes (1971) model, I focus here on a rigorous treatment of this case, which also captures the essential features of the problem for g-HawkesE(k) models. Furthermore, assuming that the deterministic functions $\mu_m(t)$ are known avoids confining the exposition to a particular parametric specification of these functions. A parametric statistical model or family of DGPs, $\{\mathsf{P}_\theta\}_{\theta\in\Theta}$, is said to be *identified* here if and only if,

$$\mathsf{P}_\theta = \mathsf{P}_{\bar{\theta}} \Rightarrow \theta = \bar{\theta} \quad \forall \theta, \bar{\theta} \in \Theta. \tag{8}$$

That is, different parameter vectors give rise to different DGPs and hence are not observationally equivalent. It is shown below that the BV linear Hawkes (1971) model is identified for any finite k, no matter how large, essentially because different parameter vectors always give rise to 'response functions' that cross only at a finite number of points. This result is stated formally in the following lemma whose proof relies on the non-singularity of generalised Vandermonde matrices (see Norberg 2002, p.2).

Lemma 2.2 *Define the response functions* $h(s;\theta) = \sum_{j=1}^{K} \alpha_j e^{-\beta_j s}$ *and* $h(s;\bar{\theta}) = \sum_{j=1}^{K} \bar{\alpha}_j e^{-\bar{\beta}_j s}$, *and the set of points where the 2 functions cross,* $C_h(\theta,\bar{\theta}) = \{s > 0 : h(s;\theta) = h(s;\bar{\theta})\}$. *Suppose that at least one element of the parameter vectors differs, that is,* $\theta \neq \bar{\theta}$. *Then, under the conditions*

$$\alpha_j, \bar{\alpha}_j > 0 \ \forall j, \ \beta_1 > \beta_2 > ... > \beta_K > 0, \ \text{and } \bar{\beta}_1 > \bar{\beta}_2 > ... > \bar{\beta}_K > 0, \tag{9}$$

the number of distinct crossing points $\#C_h(\theta,\bar{\theta}) < 2K - b$, *where* $b = \#[\{\beta_j\}_{j=1,...,K} \cap \{\bar{\beta}_j\}_{j=1,...,K}]$, $0 \leq b \leq K$, *and* $\#$ *denotes the cardinality of a finite set.*

Proof. *Given in Appendix A.* ■

Lemma 2.2 states that when at least one element of the two parameter vectors differs the number of crossing points is always strictly less than $2K$, and is central to the proof of the following identification theorem for the BV linear Hawkes (1971) model.

Theorem 2.3 ***(Identification)*** *Consider the BV linear Hawkes (1971) model with unknown parameter vector* $\theta = (\alpha_{11}^{(j)}, \beta_{11}^{(j)}, \alpha_{12}^{(j)}, \beta_{12}^{(j)}, \alpha_{21}^{(j)}, \beta_{21}^{(j)}, \alpha_{22}^{(j)}, \beta_{22}^{(j)})_{j=1,...,k}$*, and deterministic functions* $\mu_m(t)$ *assumed known for* $m = 1, 2$*. Conditions (C1) to (C3) below are sufficient for the model to be identified:*

(C1) $\{\alpha_{mr}^{(j)} > 0, \forall mr \text{ and } \forall j\}$;

(C2) $\{\beta_{mr}^{(1)} > \beta_{mr}^{(2)} > ... > \beta_{mr}^{(k)} > 0, \forall mr\}$;

(C3) $[\{\beta_{11}^{(j)}\}_{j=1,...,k} \cap \{\beta_{21}^{(j)}\}_{j=1,...,k}] = [\{\beta_{22}^{(j)}\}_{j=1,...,k} \cap \{\beta_{12}^{(j)}\}_{j=1,...,k}] = \emptyset$.

Proof. *Given in Appendix A.* ■

Condition (C3) states that for any parameter vector θ, the beta-type parameters governing the responses to type r events are all non-equal, for both $r = 1$ and $r = 2$.

2.3 Broadening the g-Hawkes model class

This section considers extending the g-Hawkes class to include non-linear intensity processes with the feature that events can exert a negative or inhibitory effect on a conditional intensity. The implications for the stationarity of the PP are derived. As we shall see, the specification testing methods developed in Section 4 are applicable to the entire g-Hawkes class and beyond.

The constraints ($\alpha_{mr}^{(j)} \geq 0, \pi_{mr}^{(j)} \geq 0 \ \forall mr, \forall j$) in Definition 2 ensure that the intensities of the BV g-HawkesE(k) process are, as they must be, strictly positive. Such constraints may appear to limit the applicability of a Hawkes approach. For example, it is sometimes important to allow negative jumps ($\alpha_{mq}^{(j)} < 0$) in the mth conditional intensity to occur in response to type q events (or more broadly in response to the value of a covariate associated with some event).

Two new non-linear g-Hawkes models are now proposed which allow for such inhibitory events by removing the constraints $\alpha_{mr}^{(j)} \geq 0$. For expositional simplicity models for a single trading day are considered. Spillover effects between days are straightforward to incorporate. First recall that the linear Hawkes term $\breve{\lambda}_m(t)$ is defined by equation (5). The Exponential model, called E-HawkesE(k) below, sets

$$\lambda_m(t) = \mu_m(t) + \exp\{\breve{\lambda}_m(t)\}, \quad m = 1, 2, \tag{10}$$

where the exponential transformation ensures positivity of the conditional intensity. The Threshold effect model, called T-HawkesE(k), has

$$\lambda_m(t) = \mu_m(t) + \max\{\epsilon_m - \mu_m(t), \breve{\lambda}_m(t)\}, \quad \epsilon_m > 0, \quad m = 1, 2. \tag{11}$$

This ensures that $\lambda_m(t) \geq \epsilon_m > 0$ for all t. The motivation is that once the financial market intensity for type m events has fallen to the threshold level ϵ_m, the occurrence of otherwise inhibitory events has no (immediate) impact.[15] The threshold ϵ_m could in principle be estimated or else set to a very small value. Setting $\epsilon_m = 0$ is not allowed since this would imply no further events of type m occurring in the future which is undesirable in this context (and would render the proposed specification testing methods of Section 4 inapplicable – see equation (21)).

Thus one can describe a broad class of g-Hawkes models, linear and non-linear, that has widespread applicability in the modelling of multivariate financial market event data. This is a convenient juncture to consider the stationarity of multivariate Hawkes processes. To do this it is natural to consider PPs on $(-\infty, +\infty)$ which have an infinite history at any time, t.[16] As in Brémaud and Massoulié (1996), I will consider both linear and non-linear HawkesE(k) processes of the form

$$\lambda_m(t) = f_m\{\breve{\lambda}_m(t)\}, \quad m = 1, 2, \tag{12}$$

where the functions $f_m : \mathsf{R} \to (0, \infty)$ and $\breve{\lambda}_m(t)$ is given as in (5) except that the range of all of the integrals there is now taken to be $(-\infty, t)$. Suitable choices of the f_m functions then yield the BV linear Hawkes (1971), E- and T- HawkesE(k) processes defined previously. Since non-constant deterministic components induce non-stationarity, $\mu_m(t) = \mu_m > 0$ will always be imposed. The following theorem gives conditions under which the processes are stationary.

Theorem 2.4 ***(Stationarity)*** *Let N be a bivariate HawkesE(k) process on $(-\infty, +\infty)$, i.e. let $N(t)$ have $\left(\mathsf{P}, \mathcal{F}_t^N\right)$-intensity given by (12). Define the 2×2 matrix A with elements $a_{mr} = \sum_{j=1}^{k} \alpha_{mr}^{(j)}/\beta_{mr}^{(j)}$. Recall that the spectral radius of A, $\rho(A)$, is the maximum of the absolute values of its eigenvalues. The following statements hold under P :*

(1) Let $f_m(x) = \mu_m + x$; if $\rho(A) < 1$, then N is a stationary, linear Hawkes (1971) process. (2) Let $f_m(x) = \mu_m + \exp(x)$ for $x \leq \log(\Lambda_m - \mu_m)$ and $f_m(x) = \Lambda_m > \mu_m$ for $x \geq \log(\Lambda_m - \mu_m)$. Then N is a stationary E-HawkesE(k) process. (3) Let $f_m(x) = \mu_m + \max\{(\epsilon_m - \mu_m), x\}, \epsilon_m > 0$; if $\rho(A) < 1$, then N is a stationary T-HawkesE(k) process.

Proof. *The function $f_m : \mathsf{R} \to (0, \infty)$ is said to be δ-Lipschitz iff $|f_m(y) - f_m(x)| \leq \delta|y - x| \;\; \forall x, y \in \mathsf{R}$,*

[15] An alternative specification would exclude *all* impact, immediate and future, of events occurring when $\lambda_m(t) = \epsilon_m$.

[16] The M-variate PP, N, is stationary iff for every $r = 1, 2, ...$and all bounded Borel subsets $B_1, ..., B_r$ of the real line, the joint distribution of $[N(B_1 + t), ..., N(B_r + t)]$ does not depend on t $(-\infty < t < \infty)$. Here the M-vector $N(B)$ gives, for $m = 1, ..., M$, the number of type m events whose occurrence time lies in the subset B.

$\delta > 0$. Recall also that $\int_0^\infty \alpha_{mr}^{(j)} e^{-\beta_{mr}^{(j)}(t)} dt = \alpha_{mr}^{(j)}/\beta_{mr}^{(j)}$. (1) Follows from Brémaud and Massoulié (1996, Theorem 7) since, trivially, f_m is δ-Lipschitz for $\delta = 1$. (2) Follows from Brémaud and Massoulié (1996, Theorem 8) since f_m is δ-Lipschitz for $\delta = (\Lambda_m - \mu_m)$, each f_m is bounded by $\Lambda = \max_m\{\Lambda_m\}$ and $\int_0^\infty t\alpha_{mr}^{(j)} e^{-\beta_{mr}^{(j)}(t)} dt = \alpha_{mr}^{(j)}/(\beta_{mr}^{(j)})^2 < \infty$. Note that $(\Lambda_m - \mu_m)$ is the derivative from the left of $f_m(x)$ at $x = \log(\Lambda_m - \mu_m)$. (3) Follows from Brémaud and Massoulié (1996, Theorem 7) since f_m is δ-Lipschitz for $\delta = 1$. ■

Note that in the E-HawkesE(k) process of Theorem 2.4 an upper bound, Λ_m, is imposed on f_m and hence on $\lambda_m(t)$.[17] This is reasonable here given the physical order processing constraints on any financial exchange, electronic or otherwise. It is also possible to show that in cases (1), (2) and (3), roughly speaking, 'the corresponding PP with no events in $(-\infty, 0)$ but the same specification of intensity dynamics on $[0, \infty)$ converges asymptotically to the stationary PP of Theorem 2.4' (see Brémaud and Massoulié 1996, Definition 1) for the exact convergence concept of *stability in variation*). The extension to the general multivariate case is a straightforward restatement of Theorem 2.4 and is left to the reader.

Conditions for the stationarity of (unrestricted) g-HawkesE(k) processes which have spillovers between days cannot be established using these methods – the additive effect on the intensity at time t of an event $T_i^{(m)}$ on a previous day now usually depends not only on $(t - T_i^{(m)})$ but also on t itself due to the fixed endpoints of trading days, τ_d. More generally, shifting the time origin, say, s time units alters the distance from a given time t to the end of the previous day.

2.4 Conditioning on covariates

An important strength of the approach to model specification, estimation and specification testing described by this paper is that it is straightforward to use the same methods when the conditioning is on $\{\mathcal{F}_t\}$, a wider filtration than the internal history.[18] Important additional information that might be included in $\{\mathcal{F}_t\}$ in this context is the covariates or 'marks' of the events other than the event type (for example, the size and direction of trades).

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 its marks. For example, consider the specification of the term $\tilde{\lambda}_{12}^{(j)}(t)$ in equations (3) and (4), and suppose that type 1 (resp. type 2) events are mid-quote

[17] The exponential feature of f_m requires this if stationarity is to be proved using these methods.
[18] The econometrician then specifies the $\mathcal{F}_t$-intensity of $N_m(t)$ for $m = 1, ..., M$.

changes (resp. trades). In (4), 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

$$\tilde{\lambda}_{12}^{(j)}(t) = \pi_{12}^{(j)}\tilde{\lambda}_{12}^{(j)}(\tau_{d-1})\mathrm{e}^{-\rho_{12}^{(j)}(t-\tau_{d-1})} + \sum_{i:T_i^{(2)}\in[\tau_{d-1},t)} Y_i\mathrm{e}^{-\beta_{12}^{(j)}(t-T_i^{(2)})}, \tag{13}$$

for $\tau_{d-1} < t \leq \tau_d$ $(d = 1, 2, ...)$, where $Y_i = g(Z_i^{(2)})$ for some parametrised function g. Such an extension should prove useful in future market microstructure research.[19]

3 Likelihood Inference

Point process models such as the g-Hawkes class, that are specified via the conditional intensity, are readily amenable to likelihood-based statistical analysis. The specification of a statistical model for multivariate PP data via a parametric family of such $\mathcal{F}_t^N$-conditional intensities is now considered before turning to the computation and properties of MLEs for g-HawkesE(k) models.

3.1 Conditional intensities and likelihoods

Theorem 3.1 below establishes that one can work directly with the parametrised family of $\mathcal{F}_t^N$-intensities, $\{\lambda_\theta(t)\}_{\theta\in\Theta}$. This family completely specifies the parametric statistical model in the sense that each $\lambda_\theta(t)$ fully determines the corresponding DGP, P_θ. Let $\lambda_\theta(t) = \{\lambda_m(t;\theta)\}_{m=1}^M$. The intuition for the result is that, since the PP is simple, each intensity $\lambda_m(t+;\theta)$ gives the conditional probability per unit time of a type m event as the time interval tends to zero (c.f. equation (38)). Furthermore, the likelihood function can be expressed in terms of the intensities, as in equation (15) below. Note that analytic likelihoods are available for the g-HawkesE(k) models since these possess intensities whose sample paths can be integrated analytically.

Theorem 3.1 ***(Likelihood Function)*** *Suppose that a parametric family of vector processes, $\{\lambda_\theta(t)\}_{\theta\in\Theta}$, is specified and let $N(t) = (N_m(t))_{m=1}^M$ be an M-variate point process on $(0, T]$, $0 < T < \infty$. Also let each $\lambda_\theta(t)$ be an M-variate, positive, $\mathcal{F}_t^N$-predictable process. Then there exists a corresponding family of DGPs, $\{\mathsf{P}_\theta\}_{\theta\in\Theta}$, such that the M-variate PP $N(t)$ has $\mathcal{F}_t^N$-intensity $\lambda_\theta(t)$ under the DGP P_θ (for all*

[19] The non-linear E- and T-HawkesE(k) models are easily adapted to this setting in order to allow Y_i to take negative values.

θ). Furthermore, each P_θ is unique in the sense that if $\lambda_\theta(t)$ is also a $(\tilde{\mathsf{P}}, \mathcal{F}_t^N)$-intensity, then P_θ equals $\tilde{\mathsf{P}}$ on $\mathcal{F}_\infty^N$.

Furthermore, suppose that $N(t)$ is a standard multivariate Poisson process under P_o, i.e. $N_m(t)$ has $(\mathsf{P}_o, \mathcal{F}_t^N)$-intensity equal to 1 for $m = 1, ..., M$, and that

$$\sum_{m=1}^{M} \int_0^T \lambda_m(s;\theta)ds < \infty \quad \mathsf{P}_o\text{-}a.s. \tag{14}$$

Then the family $\{\mathsf{P}_\theta\}_{\theta\in\Theta}$ is dominated by P_o and the likelihood function (i.e. the density of P_θ w.r.t. P_o) is given by [20]

$$L(\theta; \{N(t)\}_{t\in(0,T]}) = \exp\left\{ \sum_{m=1}^{M} \left[\int_0^T (1 - \lambda_m(s;\theta))ds + \int\limits_{(0,T]} \log \lambda_m(s;\theta)dN_m(s) \right] \right\}. \tag{15}$$

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 the P_θ corresponding to each $\lambda_\theta(t)$, see Brémaud (1981, Theorem T8, p.64).*

■

In the case where the intensities, $\lambda_m(t;\theta)$, do not have parameters in common, equation (15) yields the following log-likelihood for an M-variate PP model

$$l(\theta) = \sum_{m=1}^{M} l_m(\theta_m). \tag{16}$$

Here $\theta = (\theta_1, ..., \theta_M)$ and θ_m is the parameter vector of the intensity for type m events which varies in the parameter space Θ_m. Provided that the θ_m are variation free (i.e. θ may take any value in $\Theta_1 \times ... \times \Theta_M$), computation of the MLE can be performed via the separate maximisation of each component, $l_m(\theta_m)$, of the log-likelihood function. This is of some importance since it greatly facilitates the application of the modelling approach discussed in the paper to higher dimensional cases involving a larger number of different types of event, M.

In the case of the BV g-HawkesE(k) model given by Definition 2,

$$l_m(\theta_m) = \sum_{d=1}^{T/l} \left\{ \int_{A_d} (1 - \lambda_m(s;\theta_m))ds + \int\limits_{A_d} \log \lambda_m(s;\theta_m)dN_m(s) \right\}, \tag{17}$$

where $\lambda_m(s;\theta_m)$ is given by (3) and (4), $A_d = (\tau_{d-1}, \tau_d]$, 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

[20] Note that since (15) is a density with respect to P_o, the corresponding log-likelihood function contains the additive constant MT. Clearly, this constant does not affect the MLE.

in (4) in order to compute the log-likelihood more efficiently. As was mentioned earlier, integration of the sample path of the intensity (with respect to Lebesgue measure) can be performed analytically in the case of g-HawkesE(k) models. Evaluating (17) thus yields

$$
\begin{aligned}
l_m(\theta_m) \quad = \quad & T - \int_0^T \mu_m(s;\theta_m)ds + \sum_{d=1}^{T/l} \sum_{T_i^{(m)} \in A_d} \{\log \lambda_m(T_i^{(m)};\theta_m)\} \qquad (18) \\
& - \sum_{s=1}^{2} \sum_{d=1}^{T/l} \sum_{j=1}^{k} \left\{ \frac{\pi_{ms}^{(j)}}{\rho_{ms}^{(j)}} (1 - \mathrm{e}^{-l\rho_{ms}^{(j)}}) \tilde{\lambda}_{ms}^{(j)}(\tau_{d-1};\theta_m) + \right. \\
& \left. \sum_{\tau_{d-1} \le T_i^{(s)} < \tau_d} \frac{\alpha_{ms}^{(j)}}{\beta_{ms}^{(j)}} (1 - \mathrm{e}^{-\beta_{ms}^{(j)}(\tau_d - T_i^{(s)})}) \right\}.
\end{aligned}
$$

3.2 Maximum likelihood estimation

Ogata (1978) establishes under certain regularity conditions that the MLE for a simple, stationary, univariate PP model is consistent and asymptotically normal as $T \to \infty$, and that the likelihood ratio (LR) test of a simple null hypothesis possesses the standard χ^2 asymptotic null distribution. Denoting the asymptotic covariance matrix of the MLE by $I(\theta^*)^{-1}$, where θ^* is the true parameter value, Ogata (1978) also demonstrates (under the same regularity conditions) that

$$-\mathsf{E}_{\theta^*}\left[(1/T)(\partial^2 l(\theta^*)/\partial\theta\partial\theta')\right] \to I(\theta^*), \qquad (19)$$

where $l(\theta)$ denotes the exact log-likelihood for observation over $(0, T]$ of the stationary process (which is assumed to be a PP on $(-\infty, +\infty)$). Further details concerning these results may be found in Bowsher (2002). Section 2.3 derived conditions for the stationarity of multivariate HawkesE(k) processes, including two non-linear cases.

There are currently no results in the statistical or econometrics literature concerning the properties of MLEs for multivariate PP models. Furthermore, multivariate PP models of financial market event data are usually nonstationary owing to the presence of deterministic components in the intensities which capture intradaily seasonality. Section 6 addresses this problem via simulation, demonstrating that the MLEs and associated asymmetric confidence intervals (CIs) for the bivariate g-HawkesE(k) model are well behaved in the context of a DGP based on the estimated empirical model presented later in Section 5. The asymmetric CIs are computed here by parametrising the log-likelihood in terms of $\varphi_m = \log(\theta_m)$ ($m = 1, 2$), calculating CIs for the elements of φ_m using the inverse of the negative Hessian matrix and assuming normality in the usual manner, and then exponentiating their endpoints to obtain CIs for the elements of θ_m. The validity of this procedure is discussed further in Section 6.

Thus far I have advocated an intensity-based approach to the specification of parametric, multivariate PP models and discussed likelihood inference for such models. The development of specification testing procedures for these models based on a multivariate random change of time is one of the contributions of this paper, and is considered in the following section.

4 Specification Testing

The setting is that of a general parametric model for an M-variate PP, $N(t) = (N_m(t))_{m=1}^M$, that is specified by the family of $\mathcal{F}_t$-intensities $\{\lambda_\theta(t)\}_{\theta\in\Theta}$, where $\{\mathcal{F}_t\}$ includes but is not limited to the internal history of the M-variate PP, $\{\mathcal{F}_t^N\}$. The issue is how to construct specification tests for such a PP model and what conditions any suggested procedure imposes on the model in order to ensure its validity.

Building on the work of Ogata (1988) and Russell (1999), I show that specification testing may be performed by transforming the M non-Poisson processes, $(N_m(t))_{m=1}^M$, to *independent,* unit intensity Poisson processes, using a multivariate random change of time.[21] It appears that this is the first paper to provide an in depth treatment of specification testing for multivariate PP models. The changes of time are based on the $\mathcal{F}_t$-intensities for type m events, where $\mathcal{F}_t^N \subseteq \mathcal{F}_t$ $\forall t$. This yields an independence result that is then used to construct novel omnibus specification tests for the multivariate case. So-called m-tests are also proposed that allow investigation of the specification of the individual intensities for type m events, $\lambda_m(t)$. Ogata (1988) introduced testing procedures based on the analogous random change of time in the univariate case under the name 'residual analysis', but did not consider the conditions which the PP model and the history on which the intensity is conditioned must satisfy.

Consider the M sequences of 'generalised residuals' $\{e_i^{(m)}(\theta)\}$, $m = 1, ..., M$, where

$$e_i^{(m)}(\theta) := \int_{T_i^{(m)}}^{T_{i+1}^{(m)}} \lambda_m(s;\theta)ds \qquad (i = 0, 1, ...), \tag{20}$$

the integrand is the intensity for type m events and $(T_i^{(m)}, T_{i+1}^{(m)}]$ is the duration between adjacent type m events. Russell (1999) suggests basing a set of specification tests on each of these sequences and conjectures that the results of Yashin and Arjas (1988) could be used to prove that each sequence is *i.i.d.* Exponential(1) (when θ is the true parameter vector). However, this would require $F_i^{(m)}(x) =$

[21] The usefulness of random changes of time for specification testing in PP models has recently been emphasised by Daley and Vere-Jones (2003, Ch. 7).

$\mathsf{P}_\theta[T_{i+1}^{(m)} - T_i^{(m)} \leq x|\mathcal{G}_x]$ to be absolutely continuous (with respect to Lebesgue measure), where $\mathcal{G}_x$ is the information set on which the assessment of the hazard is based and depends on x.[22] Such a condition is undesirably restrictive in this context since it does not allow $F_i^{(m)}(x)$ to exhibit jumps in response to the occurrence of type q events ($q \neq m$).

4.1 A random time change argument

It turns out that the generalised residuals $\{e_i^{(m)}(\theta)\}$ can be identified as the durations of the Poisson processes obtained from the multivariate random time change discussed above. This provides a means of proving in considerable generality the *i.i.d.* $\mathsf{Exp}(1)$ property of the $\{e_i^{(m)}(\theta)\}$ sequences and, importantly, their independence across m. The proof imposes a much weaker condition on the PP model (equation (21) below) than the absolute continuity condition of Russell (1999). Furthermore, the intensities are conditional on $\{\mathcal{F}_t\}$ which is permitted to include information additional to the internal history of the PP. This is important in financial econometrics where it is commonly required to condition also on covariates (or 'marks') associated with each of the events. Note that Theorem 4.1 requires $\{\mathcal{F}_t\}$ to be a history of $N(t)$, that is $\{\mathcal{F}_t\}$ must contain the internal history of the *multivariate* PP $N(t)$.

Theorem 4.1 ***(Multivariate Random Time Change)*** *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 $(\mathsf{P}_\theta, \mathcal{F}_t)$-intensity $\lambda_m(t;\theta)$, where $\lambda_m(t;\theta)$ satisfies*

$$\int_0^\infty \lambda_m(t;\theta)dt = \infty \quad \mathsf{P}_\theta\text{-}a.s. \tag{21}$$

Define for each m and all $t \geq 0$, the $\mathcal{F}_t$-stopping time $\tau_m(t;\theta)$ as the (unique) solution to

$$\int_0^{\tau_m(t;\theta)} \lambda_m(s;\theta)ds = t. \tag{22}$$

Then the point processes $\{\tilde{N}_m(t;\theta)\}_{m=1}^M$ defined by

$$\tilde{N}_m(t;\theta) = N_m(\tau_m(t;\theta)), \quad t \geq 0 \tag{23}$$

are independent Poisson processes with unit intensity. Furthermore the durations, $\{(\tilde{T}_{i+1}^{(m)} - \tilde{T}_i^{(m)})\}$, of each Poisson process $\tilde{N}_m(t;\theta)$ are given by

$$(\tilde{T}_{i+1}^{(m)} - \tilde{T}_i^{(m)}) = e_i^{(m)}(\theta) \qquad (i = 0, 1, ...), \tag{24}$$

[22] Further discussion is given in Bowsher (2002, p.41).

where $e_i^{(m)}(\theta)$ is the generalised residual in (20).

It is now immediate that each sequence $\{e_i^{(m)}(\theta)\}$ is *i.i.d.* $\mathsf{Exp}(1)$ under P_θ. To prove the multivariate random time change Theorem 4.1, I shall draw on the elegant and relatively accessible proof of Brown and Nair (1988).[23] The following lemma will be needed to provide the required link between their Theorem 1 and Theorem 4.1 for time-changed counting processes.

Lemma 4.2 *Let $\{\tilde{T}_i^{(m)}\}_{i\in\{1,2\dots\}}$ be the PP that has the associated counting process $\tilde{N}_m(t;\theta)$ given by (23). Then $\tilde{T}_i^{(m)} = \int_0^{T_i^{(m)}} \lambda_m(s;\theta)ds \ \forall i, \forall m$.*

Proof. *Let s satisfy $\tilde{T}_{i-1}^{(m)} \leq s < \tilde{T}_i^{(m)}$ (for some $i \geq 1$). Then $\tilde{N}_m(s) = i-1$ and $\tilde{N}_m(s) = N_m(\tau_m(s)) = i-1$ for all such s by (23). Now as $s \uparrow \tilde{T}_i^{(m)}$, $\tau_m(s) \uparrow \tau_m(\tilde{T}_i^{(m)})$ by the continuity of τ_m and $N_m(\tau_m(s)) \uparrow i-1$. Since $N_m(\tau_m(\tilde{T}_i^{(m)})) = \tilde{N}_m(\tilde{T}_i^{(m)}) = i$ by (23), the counting process N_m must jump at $\tau_m(\tilde{T}_i^{(m)})$, that is $\tau_m(\tilde{T}_i^{(m)}) = T_i^{(m)}$. It then follows from (22) that $\tilde{T}_i^{(m)} = \int_0^{T_i^{(m)}} \lambda_m(s;\theta)ds$.* ■

The proof of Theorem 4.1 now follows.

Proof. (Theorem 4.1) *Recall that $A_m(t;\theta) := \int_0^t \lambda_m(s;\theta)ds$ is known as the F_t-compensator of $N_m(t)$ and note that $A_m(t;\theta)$ has continuous sample paths here. It follows from Lemma 4.2 that $\tilde{T}_i^{(m)} = A_m(T_i^{(m)};\theta)$. Brown and Nair (1988, Theorem 1) establishes that the M point processes $\{\{A_m(T_i^{(m)};\theta)\}_{i=1}^\infty : m = 1,...,M\}$ are independent Poisson processes with unit intensity, which completes the proof.* ■

Note that the only conditions required for Theorem 4.1 are that $\{\mathcal{F}_t\}$ is a history of $N(t)$ ($\mathcal{F}_t = \mathcal{F}_t^N$ $\forall t$ is *not* required) and that (21) is satisfied. Furthermore, (21) can be shown to hold 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 for the BV g-HawkesE(k) model to satisfy (21) for $m = 1, 2$, is that

$$\mu_m(t) \geq \zeta_m > 0 \ \forall t \text{ and } m = 1, 2, \tag{25}$$

where ζ_m is some positive constant, since this implies that $\int_0^\infty \mu_m(t)dt = \infty$ ($m = 1, 2$).[24]

The changes of time for each m in (22) are based on the $\mathcal{F}_t$-intensities, $\lambda_m(t)$, $m = 1,...,M$, where $\{\mathcal{F}_t\}$ is *identical* for each m and includes the internal history of the multivariate PP, $N(t)$. This is

[23] The first multivariate random time change result for point processes is attributed to Meyer (1971), although it is stated indirectly there.

[24] The BV g-HawkesE(k) model used later to analyse the trade and quote dataset in Section 5 has $\mu_m(t)$ given by equation (27), which clearly satisfies (25).

what underlies the result in Theorem 4.1 that the time-changed Poisson processes $\{\tilde{N}_m(t;\theta)\}_{m=1}^M$ are independent across m. If instead $\lambda_m(t)$ in (22) were the $\mathcal{F}_t^{N_m}$-intensity, where $\{\mathcal{F}_t^{N_m}\}$ is the internal history of the *univariate* PP $N_m(t)$ alone, each $\tilde{N}_m(t;\theta)$ continues to be a unit intensity Poisson process but their independence across m no longer holds in general.

4.2 Constructing specification tests

The motivation for the specification tests proposed below is as follows. Imagine the generation of a sample path of $N(t)$ conditional on its history, according to the intensities $\{\lambda_m(t)\}_{m=1}^M$. Then the level of $\lambda_m(t)$ will tend to be high over durations between type m events that turn out to be short and *vice versa.* Cumulating the levels by integration of $\lambda_m(t)$ over each $(T_i^{(m)}, T_{i+1}^{(m)}]$ yields generalised residuals whose empirical distribution is close to $\mathsf{Exp}(1)$ for a long sample path. In the context of specification testing, large deviations from the $\mathsf{Exp}(1)$ distribution then indicate that the model for $\lambda_m(t)$ explains poorly the conditional generation of the data. Under correct specification, the joint distribution of type m and type q generalised residuals ($m \neq q$) also exhibits independence.

The aim is to test the null hypothesis that there exists a $\theta^* \in \Theta$ such that the DGP is P_{θ^*}. I propose two types of such specification test for the multivariate case based on the generalised residuals $\{e_i^{(m)}\}$ in (20): novel omnibus or 'o-tests', and 'm-tests' for the specification of each intensity $\lambda_m(t)$. These are considered in turn below. Russell (1999) proposed tests similar to the m-tests but did not consider their joint distribution across m. The focus here is on the multivariate case, with tests for univariate PP models being the m-tests for a single type, m. The discussion abstracts from the impact of parameter estimation for the time being.

Theorem 4.1 states that if the DGP is P_{θ^*}, in which case $\lambda_m(t;\theta^*)$ is the true $\mathcal{F}_t$-intensity for $m = 1, ..., M$, then the time-changed PPs $\{\tilde{N}_m(t;\theta^*)\}_{m=1}^M$ are *independent* Poisson processes each with unit intensity. The superposition $\tilde{N}_0(t) := \sum_{m=1}^M \tilde{N}_m(t;\theta^*)$ is then a Poisson process with intensity M.[25] The new omnibus tests are defined as tests of the *i.i.d.* $\mathsf{Exp}(1)$ property of the rescaled durations of the superposition $\tilde{N}_0(t)$. In order to compute these durations, first 'pool' the points associated with the M sequences of generalised residuals – i.e. sort the collection of points $\{\tilde{T}_i^{(m)}\}_{i,m}$ into ascending order. Then calculate the durations between the resultant points and rescale them by multiplying by M.[26] The

[25] Since the $\tilde{N}_m(t)$ processes are independent, the $\mathcal{F}_t^{\tilde{N}}$-intensity of each $\tilde{N}_m(t)$ is 1. Thus the $\mathcal{F}_t^{\tilde{N}}$-intensity of $\tilde{N}_0(t)$ is $\sum_{m=1}^M 1$, which is therefore also the $\mathcal{F}_t^{\tilde{N}_0}$-intensity of $\tilde{N}_0(t)$.

[26] 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

advantage of omnibus tests is clear when the number of types of event M is large since consideration of many test statistics (at least one for each m) may be avoided.

The proposed m-tests are complementary and allow investigation of possible sources of misspecification via the separate consideration of each of the intensities for type m events. They are defined, for a given m, as tests of the *i.i.d.* $\mathsf{Exp}(1)$ property of the single sequence of generalised residuals $\{e_i^{(m)}(\theta^*)\}$.[27] Theorem 4.1 implies that two m-tests based on, say, the qth and rth sequences respectively ($q \neq r$) are independent, thus allowing a joint test with known size to be constructed (abstracting from parameter uncertainty). The independence result of the multivariate random time change Theorem 4.1 thus plays an important role for both o- and m-tests. In both cases, the problem of specification testing is reduced to that of testing the joint hypothesis of independence and $\mathsf{Exp}(1)$ distribution for a sequence of random variables. One can either test for exponentiality directly, or test for uniformity after applying the appropriate transformation. There are numerous ways to test a joint hypothesis of this sort, only a selection of which are discussed below.

In the empirical section of the paper, three m-tests are reported for each m ($m = 1, 2$): the Box-Ljung 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 (denoted by $BL^{(m)}$ and $BL_2^{(m)}$ respectively) and the Engle and Russell (1998) test of excess dispersion ($ED^{(m)}$) based on the $\{e_i^{(m)}(\hat{\theta}_m)\}$ series; $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ here denotes the MLE. The test of excess dispersion is a test for exponentiality and is given by $\sqrt{N_m(T)}((\hat{\sigma}^2_{\hat{e}^{(m)}} - 1)/\sqrt{8})$, where $\hat{\sigma}^2_{\hat{e}^{(m)}}$ is the sample variance of the $e_i^{(m)}(\hat{\theta}_m)$'s.[28] In addition, the analogous o-tests are also reported using the series of durations obtained by pooling the points associated with $\{e_i^{(1)}(\hat{\theta}_1)\}$ and $\{e_i^{(2)}(\hat{\theta}_2)\}$ as described above.

In all cases, p-values are calculated using the asymptotic null distributions which hold in the case of known parameters. It is demonstrated using simulation in Section 6 that this 'plug-in' approach does not result in serious size distortions, at least in the context of a DGP based on the estimated empirical model presented in Section 5. There is evidence that the $ED^{(m)}$ tests based on the MLE are undersized, raising concerns about their power properties when the nominal critical values are used. The difficulties

different intervals since the change of time results in all processes having unit intensity (e.g. consider the case where $N_m(T)$ is much greater than $N_q(T)$). The following pragmatic approach to this problem is adopted later in Section 5.2. Let $\tilde{T}_*^{(m)}$ be the last observed point of $\tilde{N}_m(t)$ and $\tilde{T}_* := \min\{\tilde{T}_*^{(m)}, m = 1, ..., M\}$. Then sort the points $\{\tilde{T}_i^{(m)} : \tilde{T}_i^{(m)} \leq \tilde{T}_*, m = 1, ..., M\}$ into ascending order and proceed as before.

[27] Berman (1983), in the context of non-stationary Poisson processes, seems to have been the first to suggest testing the properties of the durations of the transformed PP in this manner in order to assess goodness-of-fit.

[28] The test statistic would have an asymptotic $\mathsf{N}(0, 1)$ distribution if $\{e_i^{(m)}(\hat{\theta}_m)\}$ were *i.i.d.* $\mathsf{Exp}(1)$.

created by parameter estimation error in this context are now discussed, together with a partial solution.

Tests of time series model specification are often based on the *i.i.d.* $\mathsf{Uniform}(0,1)$ generalised residuals obtained from the transform based on the conditional distribution functions (known as the 'probability integral transform') – see e.g. Smith (1985), Shephard (1994), and Diebold, Gunther, and Tay (1998). Thompson (2001, 2004) and Hong and Li (2005, Section 1.2) provide critical values for joint tests of the independence and uniformity of the generalised residuals that are asymptotically exact in the presence of parameter estimation error. In the case of a univariate PP, it can be shown that the sequence $\{1 - \exp(-e_i)\}$ is identical to these *i.i.d.* $\mathsf{U}(0,1)$ generalised residuals, but no such equivalence statement holds for the $\{e_i^{(m)}\}$ in the multivariate case. One reason is that whereas in the time series setting, the conditioning information set is fixed each period, the relevant information set for the duration $S_{i+1}^{(m)} := (T_{i+1}^{(m)} - T_i^{(m)})$ is not constant over that duration.

An asymptotic theoretical treatment of the impact, in the multivariate case, of parameter estimation on the null distributions of both o- and m-tests is therefore an open, challenging problem and the results of Thompson (2001, 2004) and Hong and Li (2005, Section 1.2) do not transfer readily. A solution is to consider 'p-tests' based on the generalised residuals of the 'pooled process', $N_0(t) := \sum_{m=1}^{M} N_m(t)$, and given by

$$e_i^{(p)}(\theta) := \int_{T_i}^{T_{i+1}} \sum_{m=1}^{M} \lambda_m(s;\theta)ds \qquad (i = 0, 1, ...). \tag{26}$$

Provided that $\mathcal{F}_t = \mathcal{F}_t^N \ \forall t$, the conditioning information set $\sigma(T_1, Z_1, ..., T_i, Z_i)$ is now constant over (T_i, T_{i+1}). Since the $\mathcal{F}_t^N$-intensity of $N_0(t)$ is given by $\sum_{m=1}^{M} \lambda_m(t)$, $e_i^{(p)}(\theta^*) \sim$ *i.i.d.* $\mathsf{Exp}(1)$ and $[1 - \exp(-e_i(\theta^*))] \sim$ *i.i.d.* $\mathsf{U}(0,1)$. I conjecture that the critical values of the tests of Thompson (2004) will be asymptotically exact for such p-tests in the presence of parameter estimation error.[29] Similarly, the nonparametric test of Hong and Li (2005, Section 1.2) might also be applied in this context. The p-test solution is a partial one since the part of the model specification determining the conditional generation of the type of the next event, Z_{i+1}, is not being directly tested.

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 a linear BV g-HawkesE(k) model to the timing of trades

[29] The setting is not identical to Thompson (2004) due to the conditioning on the additional covariates $(Z_1, ..., Z_i)$.

and mid-quote changes for a NYSE stock. The specification testing procedures developed in Section 4 above 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. Engle (2002, p. 427) comments on the desirability of an "ultra high frequency" model that can "give a continuous record of instantaneous volatility, where events such as trades and quote revisions as well as time itself modify the volatility estimate." It will be shown in Section 5.3 that the BV g-HawkesE(k) model (see Definition 2) can be used to obtain just such an estimate of the instantaneous volatility of the price process.

5.1 Data and institutional background

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 ask) communicated to the trading crowd and to execute any order at a price that is at least as favourable as his published quote.[30] 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 (whose average differs from the previous mid-quote, as is usually the case). A trade occurs when the specialist matches a buy and a sell order (or alternatively, takes one side of the transaction himself). Thus, 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

[30] 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).

deleted from the dataset.[31] Having mapped the times of the market events onto $(0, \infty)$ using the data transformation described in Section 2.1 (with the day length l set to 6.5 hours), the times of the mid-quote changes were then thinned (i.e. a subset of the times was selected) as in Engle and Russell (1998) in order 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.[32] The threshold of \$1/16 used here is approximately equal to half of the average spread of \$0.117, and thus represents a quite small movement in the mid-quote. In order to avoid the modelling task becoming overly complicated, the absolute size of the change in the mid-quote between successive events is not modelled directly. Given the way the mid-quote events are defined, this can exceed \$1/16. However, analysis of this dataset showed that the majority (82.4%) of these changes do indeed equal \$1/16.[33]

Recall from the definition of an M-variate PP that, since the associated '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 measurement error – i.e. 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 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. Further discussion of the treatment of events with identical timestamps is given in Appendix C, together with the results of a sensitivity analysis that examines two other adjustment rules. These results strongly support the robustness of the main empirical findings reported below.

[Table 1 here]

[31] Trades relating to the opening auctions were also deleted, as were trades that were subsequently cancelled. 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.

[32] The first mid-quote event is defined to occur at the time when the first pair of quotes was 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.

[33] 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.

Summary statistics of the intertrade durations and durations between successive mid-quote events for the final dataset are given in Table 1.[34] 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 function for the deterministic components, $\mu_m(t)$, in (3) worked well in practice. The spline is continuous over the course of the 6.5 hour trading day, with knots 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:

$$\mu_m(t;\gamma_m) = \begin{cases} 1_{v(t)\in(0,0.5]}[\gamma_{m1} + 2v(t)(\gamma_{m2} - \gamma_{m1})] + \\ \sum_{i=1}^{6} 1_{v(t)\in(i-0.5,i+0.5]}[\gamma_{m,i+1} + (v(t) - i + 0.5)(\gamma_{m,i+2} - \gamma_{m,i+1})] & \text{for } v(t) > 0, \\ \gamma_{m8} & \text{for } v(t) = 0, \end{cases} \tag{27}$$

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.

The results of fitting the unrestricted BV g-HawkesE(2) model are not given here. Rather Table 2 reports results for a restricted model that was found to have identical log-likelihood and specification test statistics.[35] This restricted BV g-HawkesE(2) model has the restrictions ($\pi_{mm}^{(1)} = 0$, $\pi_{mq}^{(1)} = 0, \pi_{mq}^{(2)} = 0; m = 1, 2$; $q = 2$ if $m = 1$ and $q = 1$ if $m = 2$) imposed. These imply that the stochastic components of the intensity exhibit zero spillover effects, except for the $\tilde{\lambda}_{mm}^{(2)}(t)$ $(m = 1, 2)$ components in equation (3) which capture the $j = 2$ effect of type m events on the mth intensity. The spillover effects are discussed in more detail below. Results are considered only for models with $k = 2$ since such E(2) models were found to be clearly superior both in terms of specification tests and log-likelihood to E(1) models. Numerical optimisation of the log-likelihood was performed using the MaxBFGS algorithm with numerical derivatives in Ox (see Doornik 2001).[36]

[Table 2 here]

[34] 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.

[35] The 'restricted parameters' had MLEs very close to zero in the unrestricted model and the MLEs of the 'unrestricted parameters' were virtually identical for the two models.

[36] For this, a sequential estimation procedure was used involving univariate non-homogeneous Poisson, Hawkes (1971) and g-HawkesE(k) models and bivariate Hawkes (1971) and g-HawkesE(k) models. Details of this procedure and of the recursions used to improve computational efficiency may be found in Bowsher (2002).

The restricted model fits very well indeed. Only one of the m-tests of specification in Table 2 rejects at the 5% level. Furthermore, the omnibus or o-test statistics $(BL^{(o)}, BL_2^{(o)}, ED^{(o)})$ were $(23.670, 23.277, -3.7246)$, corresponding to p-values of $(0.0709, 0.0784, 0.0002)$ respectively. For discussion of the m- and o-tests used see Section 4.2. The adequacy of the specification runs counter to the view sometimes encountered that the structure of the conditional intensities of linear Hawkes models may be too restrictive in the context of financial market event data.

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 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 the 'cross effect' component $\sum_{j=1}^{2} \tilde{\lambda}_{mq}^{(j)}(t)$ in (3); and the fourth shows $\hat{\lambda}_{mm}(t)$, the estimate of the 'self exciting' component $\sum_{j=1}^{2} \tilde{\lambda}_{mm}^{(j)}(t)$, together with S_{mm}, the estimate of the spillover effect $\pi_{mm}^{(2)} \tilde{\lambda}_{mm}^{(2)}(\tau_{20}) \mathrm{e}^{-\rho_{mm}^{(2)}(t-\tau_{20})}$.

[Figure 1 here]

Consider first the MLEs reported in Table 2. Asymmetric 95% confidence intervals (CIs) are shown in square parentheses.[37] 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 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 the cross effect parameters $\alpha_{mq}^{(1)}, \beta_{mq}^{(1)}, \alpha_{mq}^{(2)}$, 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

[37] These were computed using the method described in Section 3.2. 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.

$\hat{\lambda}_{12}(t)$ 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 (3), $\sum_{j=1}^{2} \tilde{\lambda}_{mq}^{(j)}(t)$, 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 in Section 5.4 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. The standard LR test cannot be applied here since the parameter value under the null lies on the boundary of the maintained hypothesis, and there are nuisance parameters ($\pi_{12}^{(j)}, \rho_{12}^{(j)}, \beta_{12}^{(j)}; j = 1, 2$) that are identified under the alternative but not under the null.[38] Nevertheless, it is noted that imposing the restrictions $\alpha_{12}^{(j)} = 0$ ($j = 1, 2$) on the $\lambda_1(t)$ intensity of the 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 specification 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 Appendix C 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 confirms 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 linear Hawkes (1971) model with $k = 2$ (see (5)) 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 linear Hawkes (1971) model has a log-likelihood of 143,461 (corresponding to a LR test with p-value 0.000) 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

[38] The same comments apply to a test of $H_0 : \pi_{mr}^{(j)} = 0$ against the alternative $H_1 : \pi_{mr}^{(j)} > 0$ (in which case $\rho_{mr}^{(j)}$ is unidentified under the null). 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.

(a reduction of 2) and very similar MLEs and specification test statistics to those presented in Table 2.[39] In the absence of a likelihood ratio 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 (as evidenced by the specification tests obtained under the restrictions $\pi_{mr}^{(j)} = 0$). This is perhaps surprising and merits further investigation.

Elsewhere, I present evidence for the empirical importance of positive spillover effects in g-HawkesE(k) models of the timing of trades for NASDAQ stocks (see Bowsher (2002, Table 2)). The absence of such positive spillover effects here may well be due to the different opening procedures on NASDAQ and the NYSE. Whereas normal, 'continuous' trading simply recommences each day on NASDAQ after a preopening period in which virtually no trade occurs, data relating to the NYSE opening auction is omitted from the analysis in the present study. It would be interesting in future work to model jointly the opening auction and the continuous trading process on the NYSE. A general framework has been proposed here for modelling dependence between trading days in PP models of financial markets. A detailed empirical examination of spillover effects in different settings would be worthwhile, but is beyond the scope of this paper.

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 $P(t) = P(0) + \sum_{i:T_i^{(1)} \leq t} \Delta_i$. Also define the instantaneous conditional volatility by

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

[39] The MLEs and specification tests for this model are presented in Tables 3 and 4 as part of the simulation study in Section 6.

where $\mathcal{F}_t = \sigma(P(t)) \vee \mathcal{F}_t^N$ and N is the bivariate PP of trades and mid-quote events.[40] It is then possible to express $\sigma^2(t)$ in terms of the mid-quote intensity. Specifically,

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

where 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). Engle and Russell (1998) note an analogous relationship for their univariate model of price events; the difference here is that the instantaneous volatility is also conditional on the timing of trades.

[Figure 2 here]

An approximate estimate of the instantaneous volatility can be formed using the estimates for the restricted BV-g-HawkesE(2) model obtained above by making the following substitutions in (29): replace $\lambda_1(t+)$ by the (right continuous version of the) estimated intensity $\hat{\lambda}_1(t)$, set $c = \$\ 1/16$ and let $P(t) = P(0) + \sum_{i:T_i^{(1)} \leq t} \Delta_i$ (where $T_i^{(1)}$ is the time of the ith mid-quote event for the dataset and Δ_i is the actual change to the mid-quote since the last mid-quote event).[41] 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 visually as a result of the relatively small variability of the price compared to that of the intensity. The multivariate g-Hawkes models can thus be used to obtain a microstructure view of the interaction between financial market volatility and its determinants. It would be interesting in future work to condition also on the volume of trades and the direction of the price changes between mid-quote events.

[40] The smallest σ-field containing $\sigma(P(t))$ and $\mathcal{F}_t^N$ is denoted by $\sigma(P(t)) \vee \mathcal{F}_t^N$.

[41] 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, (29) holds exactly when $\Delta_i^2 = c^2\ \forall i$ but we observe $\Delta_i^2 > c^2$ for a minority of mid-quote events in the dataset. Nevertheless, the approximation is a useful one.

5.4 Connections with the market microstructure literature

Much of the existing theoretical and empirical market microstructure literature concentrates on the impact of trades on prices rather than on the reciprocal effect of prices on the trade arrival process. By contrast, the modelling framework presented here allows the two-way interaction of trades and mid-quote changes to be analysed. The empirical results of 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.[42] 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 also triggers a mid-quote change when one of the sides to the trade is 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

[42] Useful reviews of this literature are given by O'Hara (1995) and Hasbrouck (1996).

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 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. As has already been discussed, it is possible to allow for inhibitory effects using the g-Hawkes model class.

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 high volatility tend to occur in periods when there are many trades. A number of other empirical studies have reported similar findings.[43] 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.[44] 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 Monte Carlo Study of MLEs and Specification Tests

This section uses simulation to examine the appropriateness of the asymptotic approximations employed in Section 5 in the context of a DGP based on the estimated empirical model for the General Motors (GM)

[43] 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.

[44] 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.

dataset presented there. Specifically, I consider the bias of the MLEs, the coverage of the asymmetric confidence intervals (CIs) computed as described in Section 3.2 and the size properties of the o- and m-tests of specification obtained using both the true parameters and the 'plug-in' method which replaces those parameters by their MLEs.

The DGP used in the simulation study is a BV g-HawkesE(2) model over 40 trading days with no dependence across days (i.e. the restrictions $\pi_{mr}^{(j)} = 0 \; \forall mr$ and $\forall j$ apply) and with parameter values set equal to the MLEs obtained by fitting the same model to the GM dataset. This DGP was chosen because of the importance of 'independence across days' as a baseline case. As was noted earlier, the results of fitting this model to the GM dataset were very similar indeed to those presented for the slightly more general model in Table 2. The results for the no dependence model are labelled 'GM dataset' in Tables 3 and 4 below.

Computationally efficient simulation of an M-variate g-HawkesE(k) process with $\mathcal{F}_t^N$-intensity $\lambda(t) = (\lambda_m(t))_{m=1}^M$ may be performed using a modified version of the thinning algorithm of Ogata (1981). Thinning simulation algorithms for PPs in which the realised level of the 'bounding process' depends on the past of the proposal process, as well as on that of the target process have been suggested elsewhere (see e.g. Ogata (1981)), but without proof of their validity. Presented below is such an algorithm for the simulation of the BV g-HawkesE(k) process on $(0, T]$, together with the requisite proof. Here $\lambda_m(t)$ should be interpreted as the *realised* intensity for type m events of the target process at time t, conditional on the simulated past of that process during the interval $(0, t)$. For details of the general simulation approach the reader is referred to Ogata (1981) and Daley and Vere-Jones (2003, Section 7.5).

Algorithm 6.1 ***Simulation of BV g-HawkesE(k) process on*** $(0, T]$.

1. *Set* $t = 0$ *and* $d = 1$;
2. *If* $d = (T/l) + 1$, *terminate the simulation, otherwise:*
   (a) *Set* $B(t) = K + \sum_{m=1}^2 \lambda_m(t+)$, *where the constant* K *does not depend on* d *and is defined as*
   $$K := \max_{(t',t'')} \left\{ \sum_{m=1}^2 \mu_m(t'') - \sum_{m=1}^2 \mu_m(t'+) : \tau_{d-1} \le t' < t'' \le \tau_d \right\},$$
   the maximum increase in $\sum_{m=1}^2 \mu_m(t)$ *possible over any interval of time* $(t', t'']$ *contained in some arbitrary day* d, *and* $\lambda_m(t+) := \lim_{s \downarrow t} \lambda_m(s)$;
   (b) *Generate an exponential r.v.* S *with mean* $1/B(t)$, *and a uniform r.v.* U *distributed on* $(0, 1)$;
   (c) *If* $t + S > \tau_d$, *set* $t = \tau_d$, $d = d + 1$, *and return to the beginning of Step 2;*

(d) A point of the proposal process has now been generated at time $t+S$;
If $U \leq \lambda_1(t)/B(t)$, accept $t+S$ as the occurrence time of a type 1 event;
If $\lambda_1(t)/B(t) < U \leq [\lambda_1(t)+\lambda_2(t)]/B(t)$, accept $t+S$ as the occurrence time of a type 2 event;
Otherwise reject ('thin') the proposal of an event at time $t+S$;

(e) Set $t = t+S$ and return to Step 2a. ■

Proof. *Let $\lambda^*(t)$ be the process with piecewise constant, left continuous sample paths that jump to take the value $B(t)$ 'just after' time t on every occasion when t changes in Algorithm 6.1, and are constant otherwise. The processes $\lambda_1(t), \lambda_2(t)$ and $\lambda^*(t)$ are all $\mathcal{F}^*_t$-predictable where $\{\mathcal{F}^*_t\}_t := \{\mathcal{F}^N_t \vee \mathcal{F}^{N^*}_t\}_t$, which also includes the history of the proposal point process, N^*.*[45] *By Proposition 1 of Ogata (1981), Algorithm 6.1 therefore results in a DGP, P say, such that the bivariate PP has $(\mathsf{P}, \mathcal{F}^*_t)$-intensity $\lambda(t) = (\lambda_1(t), \lambda_2(t))'$. Now since $\lambda(t)$ is the $(\mathsf{P}, \mathcal{F}^*_t)$-intensity, $\mathcal{F}^N_t \subset \mathcal{F}^*_t$ $\forall t$, and $\lambda(t)$ is adapted to $\{\mathcal{F}^N_t\}$ and càglàd, then $\lambda(t)$ is also the $\mathcal{F}^N_t$-intensity under the DGP P (see Definition 3).* ■

[Table 3 here]

The main features of the results of the simulation study presented in Tables 3 and 4 are as follows. The MLEs and CIs for the elements of θ_1 and θ_2 are well behaved (see Table 3).[46] In particular, the coverages of the CIs lead us to conjecture that $\hat{\phi} := \log(\hat{\theta})$ is indeed approximately distributed as $\mathcal{N}(\phi^*, -[\partial^2 l_T(\phi^*)/\partial\phi\partial\phi']^{-1})$ for this DGP, and very likely more generally.[47] Nor does estimation by ML appear to result in large biases for the value of T used here. Turning now to the properties of the omnibus and m-tests of specification shown in Table 4 , the tests based on the true parameter vector θ^* have actual size which is well approximated by the nominal size at both the 5% and 1% levels. The use of the 'plug-in' method causes little size distortion for any of the Box Ljung tests but does result in the $ED^{(m)}$ tests becoming undersized. This raises some concerns about the power properties of the latter when the critical values based on the known parameter case are used. Examination of Table 4 suggests that the size distortions are mitigated somewhat in the case of the omnibus-type $ED^{(o)}$ tests.

[Table 4 here]

[45] Note, however, that $\lambda^*(t)$ is *not* $\mathcal{F}^N_t$-predictable because $\lambda^*(t)$ is not adapted to the natural filtration of the target process, $\{\mathcal{F}^N_t\}$, as $\lambda^*(t)$ also depends on the timing of the thinned points.

[46] The CIs for θ_1 were originally computed using the method described in footnote 37. However, it was found that inversion of the Hessian then resulted in negative variances in about 30% of the replications. The CIs shown were therefore based on a log-likelihood in which $\alpha^{(2)}_{11}, \gamma_{17}$, and γ_{18} are also treated as known. It appears that $\alpha^{(2)}_{11}$ is sufficiently close to zero to cause problems in this regard.

[47] Similar coverages were observed in a simulation not reported here using a univariate g-HawkesE(2) process with strong dependence between trading days as the DGP.

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 conditional intensity. This is a powerful and natural approach to specification, since one considers how the intensity for each type of event changes as the information set is updated in continuous time. Furthermore, such an intensity-based approach readily lends itself to both likelihood inference and specification testing based on a multivariate random time change.

The new class of generalised Hawkes models is introduced. This includes non-linear models that allow for inhibitory events resulting in a decrease in a conditional intensity and models that allow for dependence between trading days. A recursive model structure is adopted in which the stochastic components of the intensity on trading day d are specified in terms of functionals of their paths on previous days and the contributions of the events occurring on day d. The structure takes into account the existence of non-trading periods and provides a useful framework for approaching the issue of dependence between trading days. A computationally efficient thinning algorithm for g-HawkesE(k) processes is developed and is employed in a Monte Carlo study demonstrating that the asymptotic approximations used for inference in the paper are well behaved. A new result concerning identification of the bivariate linear Hawkes (1971) model is also proven.

The paper provides what is ostensibly the first in depth analysis of specification testing for parametric, multivariate point process models. Novel omnibus, or o-tests for the multivariate case are proposed together with m-tests of the specification of each of the intensities for type m events. The tests are based on a multivariate random time change that transforms the non-Poisson processes into independent Poisson processes. Both types of test are joint tests of the independence and exponential distribution of durations obtained from these Poisson processes. The technique is shown to have widespread applicability since the restriction it imposes on the model is natural in the context of financial market event data.

A full bivariate point process 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 also mid-quote events result in increased trade intensity. The estimated mid-quote intensity is used to provide an approximation to the instantaneous price volatility of the stock, in which events such as trades and quote revisions as well as time itself modify the volatility estimate as time evolves continuously.

APPENDIX

A Proofs

Presented below are the proofs pertaining to identification of the BV linear Hawkes (1971) model.

Proof. (Lemma 2.2) The more difficult case $0 < b < K$ is given in order to illustrate the method. Denote the b common parameters by $\{\beta_j^C\}_{j=1,...,b}$, and the remainder by $\{\beta_j^D\}_{j=1,...,K-b}$ and $\{\bar{\beta}_j^D\}_{j=1,...,K-b}$ for $h(s;\theta)$ and $h(s;\bar{\theta})$ resp. Suppose that $\#C_h(\theta,\bar{\theta}) = 2K - b$ and write $C_h(\theta,\bar{\theta}) = \{s_i\}_{i=1,...,2K-b}$. Define the $(2K-b) \times (2K-b)$ generalised Vandermode matrix

$$V(\theta,\bar{\theta}) := \{e^{-\beta_1^D s_i}, ..., e^{-\beta_{K-b}^D s_i}, e^{-\bar{\beta}_1^D s_i}, ..., e^{-\bar{\beta}_{K-b}^D s_i}, e^{-\beta_1^C s_i}, ..., e^{-\beta_b^C s_i}\}_{i=1}^{2K-b}, \tag{30}$$

where the ith row has no 2 elements equal $(i = 1, ..., 2K - b)$. It follows from the Theorem of Norberg (2002, p.2) that $V(\theta,\bar{\theta})$ is non-singular. Also define, using an obvious notation,

$$\alpha(\theta,\bar{\theta}) := (\alpha_1^D, ..., \alpha_{K-b}^D, -\bar{\alpha}_1^D, ..., -\bar{\alpha}_{K-b}^D, \alpha_1^C - \bar{\alpha}_1^C, ..., \alpha_b^C - \bar{\alpha}_b^C)'. \tag{31}$$

Then $\#C_h(\theta,\bar{\theta}) = 2K - b$ implies that

$$V(\theta,\bar{\theta})\alpha(\theta,\bar{\theta}) = 0, \tag{32}$$

which yields a contradiction since, by (9), $\alpha(\theta,\bar{\theta}) \neq 0$. It follows therefore that $\#C_h(\theta,\bar{\theta}) < 2K - b$. ■

Proof. (Theorem 2.3) Let $\lambda_0(t;\theta) := \sum_{m=1}^{M} \lambda_m(t;\theta)$ be the $(\mathsf{P}_\theta, \mathcal{F}_t^N)$-intensity of the pooled process $\{T_i\}$,with $\lambda_0(t;\bar{\theta})$ defined analogously as its $(\mathsf{P}_{\bar{\theta}}, \mathcal{F}_t^N)$-intensity. Then the 'uniqueness' of predictable intensities (see Brémaud 1981, Theorem T12, p.31) gives

$$\mathsf{P}_\theta = \mathsf{P}_{\bar{\theta}} \Rightarrow \lambda_0(T_2;\theta) = \lambda_0(T_2;\bar{\theta}) \quad \mathsf{P}_\theta\text{-a.s.} \tag{33}$$

I prove that the RHS of (33) does not hold when $\theta \neq \bar{\theta}$, thus establishing that $\theta \neq \bar{\theta} \Rightarrow \mathsf{P}_\theta \neq \mathsf{P}_{\bar{\theta}}$.

Define $h_r(s;\theta_r) = \sum_{m=1}^{2}\sum_{j=1}^{k}\alpha_{mr}^{(j)}\mathrm{e}^{-\beta_{mr}^{(j)}s}$ for $r = 1, 2$. Now suppose at least one element of θ and $\bar{\theta}$ differs, and say w.l.o.g. that we have $\theta_1 \neq \bar{\theta}_1$. Then on the event $\{Z_1 = 1\}$, $\lambda_0(T_2;\theta) = \lambda_0(T_2;\bar{\theta})$ iff $h_1(T_2-T_1;\theta_1) = h_1(T_2-T_1;\bar{\theta}_1)$, which by Lemma 2.2 and (C1), (C2) and (C3) holds for at most $(4k-1)$ distinct values of the duration $(T_2 - T_1)$. This follows since application of Lemma 2.2 here with $K = 2k$ gives $\#C_{h_1}(\theta_1, \bar{\theta}_1) < 4k$.

Now, denoting the event $[\lambda_0(T_2;\theta) = \lambda_0(T_2;\bar{\theta})]$ by E, $\mathsf{P}_\theta[E] = \mathsf{P}_\theta[\{E\cap\{Z_1 = 1\}\}\cup\{E\cap\{Z_1 = 2\}\}]$. Since $\{E\cap\{Z_1 = 1\}\} = \{(T_2 - T_1) \in C_{h_1}(\theta_1, \bar{\theta}_1)\}$, the duration $(T_2 - T_1)$ is a continuous r.v. $\forall\theta$, and $C_{h_1}(\theta_1, \bar{\theta}_1)$ is a finite set, it follows that $\mathsf{P}_\theta[E\cap\{Z_1 = 1\}] = 0$. Furthermore, $\mathsf{P}_\theta[E\cap\{Z_1 = 2\}] \leq \mathsf{P}_\theta[Z_1 = 2] < 1$ since $\mu_1(t) > 0\ \forall t$. Therefore

$$\mathsf{P}_\theta[\lambda_0(T_2;\theta) = \lambda_0(T_2;\bar{\theta})] = 0 + \mathsf{P}_\theta[E\cap\{Z_1 = 2\}] < 1, \tag{34}$$

and by (33) it follows that $\mathsf{P}_\theta \neq \mathsf{P}_{\bar{\theta}}$. The case where $\theta_2 \neq \bar{\theta}_2$ is symmetric with the one just given. Hence $\theta \neq \bar{\theta}$ implies that $\mathsf{P}_\theta \neq \mathsf{P}_{\bar{\theta}}$, the condition required for identification. ■

B A Primer on Point Process Theory

Textbook treatments of the martingale-based, intensity theory of PPs are given in Brémaud (1981) and Karr (1991). Since these accounts are rather demanding mathematically, and so that the paper is self-contained, this appendix summarises the aspects of the theory that are needed here.

B.1 Point and counting processes

A *simple Point Process on* $(0,\infty)$, denoted by $\{T_i\}_{i\in\{1,2,...\}}$, is a sequence of positive random variables with the property that the realisations of the sequence are strictly increasing (almost surely). Let $\{Z_i\}_{i\in\{1,2,...\}}$ be a sequence of $\{1, 2, ..., M\}$-valued random variables ($1 \leq M < \infty$). Then, if $\{T_i\}$ is a simple PP, the double sequence $\{T_i, Z_i\}$ is called an *M-variate point process on* $(0,\infty)$. Define for all $m \in \{1, ..., M\}$ and all $t \geq 0$

$$N_m(t) = \sum_{i\geq 1} 1(T_i \leq t)1(Z_i = m). \tag{35}$$

Then the process $N(t) := (N_m(t))_{m=1}^{M}$ is the *M-vector counting process associated with* $\{T_i, Z_i\}$. The ith event is said to be *of type* m (for a particular realisation) if and only if $Z_i(\omega) = m$.

The (stochastic Stieltjes) integral of a process X with respect to the counting process N_m is defined pathwise and, where it exists, is given by

$$\int_{(0,t]} X(s)dN_m(s) = \sum_{i\geq 1} 1(T_i^{(m)} \leq t)X(T_i^{(m)}), \quad \forall t > 0. \tag{36}$$

The *natural filtration* (or *internal history*) of the M-variate PP $N(t)$ is denoted by $\{\mathcal{F}_t^N\}_{t\geq 0}$, where $\mathcal{F}_t^N = \sigma(N_A(s) : 0 \leq s \leq t,\, A \in \mathcal{E})$, $N_A(s) = \sum_{i\geq 1} 1(T_i \leq s)1(Z_i \in A)$, and $\mathcal{E}$ is the σ-field of all subsets of $\{1, 2, ..., M\}$.

B.2 Conditional intensities

The martingale-based definition of the conditional intensity of a simple PP is unfamiliar in econometrics, but leads to an extremely powerful martingale calculus for *dynamic* point process systems such as those encountered in financial econometrics (see Brémaud (1981)).

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

$$\mathsf{E}[N(s) - N(t)|\mathcal{F}_t] = \mathsf{E}\left[\int_t^s \lambda(u)du|\mathcal{F}_t\right] \quad \mathsf{P}\text{-}a.s., \tag{37}$$

for all t, s such that $0 \leq t \leq s$, then $\lambda(t)$ is the $(\mathsf{P}, \mathcal{F}_t)$-intensity of $N(t)$.[48]

Alternatively, we might then say that $\lambda(t)$ is the $\mathcal{F}_t$-conditional intensity of $N(t)$ under the DGP P. In the multivariate case ($M > 1$), the now M-variate process $\lambda(t) = (\lambda_m(t))_{m=1}^M$ is the $(\mathsf{P}, \mathcal{F}_t)$-intensity of $N(t) = (N_m(t))_{m=1}^M$ if and only if $\lambda_m(t)$ is the $(\mathsf{P}, \mathcal{F}_t)$-intensity of $N_m(t)$ for $m = 1, ..., M$.

An intuitive understanding of an intensity can be gained by letting $s \downarrow t$ in (37) above. Suppose that, for each m, $N_m(t)$ is observed on $[0, T]$ and has the $(\mathsf{P}, \mathcal{F}_t)$-intensity $\lambda_m(t)$, that the sample paths of $\lambda_m(t)$ are *càglàd* and that $\{\lambda_m(t)\}_{0\leq t\leq T}$ is bounded by a P-integrable r.v. Define $\lambda_m(t+) := \lim_{s\downarrow t} \lambda_m(s)$. Then

$$\lim_{s\downarrow t} \frac{1}{s-t} \mathsf{E}[N_m(s) - N_m(t)|\mathcal{F}_t] = \lambda_m(t+) \ \ \mathsf{P}\text{-a.s.}, \tag{38}$$

(see, for example, Aalen 1978, Lemma 3.3).[49]

C 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) – that is, they are simple PPs. This section describes the

[48] Note that sufficient conditions for $\lambda(t)$ to be $\mathcal{F}_t$-predictable are that the sample paths of the process are left continuous and have right-hand limits (i.e. the paths are *càglàd*) , and that $\lambda(t)$ is adapted to $\{\mathcal{F}_t\}$.

[49] Intuitively, it is the right continuous process $\lambda_m(t+)$ rather than $\lambda_m(t)$ that appears in (38), because if there is a jump at time t then $\int_t^{t+dt} \lambda_m(u)du \simeq \lambda_m(t+)dt \neq \lambda_m(t)dt$.

adjustments that were made to the General Motors (GM) dataset analysed in Section 5 in order to deal with such simultaneous events. The timestamps of all events in the original dataset are an integer number of seconds. 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. The occurrence of such events was rare.[50] The treatment of the simultaneous occurrence of trades and mid-quote events, which arises largely as the result of measurement error, is a more substantive issue. Approximately 11 per cent of the mid-quote events have exactly the same timestamp in the data 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 of the mid-quote event and the trade are the same, the actual occurrence times (in continuous time) will rarely be identical. 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. For example, 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 being reported with identical timestamps as a result of lengthier reporting delays.[51]

Since it would be expected that the estimates of $\alpha^{(1)}_{mq}, \beta^{(1)}_{mq}, \alpha^{(2)}_{mq}, \beta^{(2)}_{mq}$ $(m = 1, 2; q \neq m)$ reported in Table 2 are affected by 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

[50] 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.

[51] 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.

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 $\alpha^{(1)}_{mq}, \beta^{(1)}_{mq}, \alpha^{(2)}_{mq}, \beta^{(2)}_{mq}$ $(m = 1, 2; q \neq m)$, and resulted in specification tests that accepted at the 1% level with the exception of the $BL^{(1)}$ test for the trade intensity. The estimates of $(\alpha^{(1)}_{12}, \alpha^{(2)}_{12}, \beta^{(1)}_{12}, \beta^{(2)}_{12}; \alpha^{(1)}_{21}, \alpha^{(2)}_{21}, \beta^{(1)}_{21}, \beta^{(2)}_{21})$ 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 the most novel of the empirical findings in Section 5.2. Crucially, although the estimates of $\alpha^{(1)}_{21}$ and $\alpha^{(2)}_{21}$ 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 $\alpha^{(j)}_{21} = 0$ $(j = 1, 2)$ 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.

References

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

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

Bauwens, L. and P. Giot (2001). Econometric modelling of stock market intraday activity. Advanced Studies in Theoretical and Applied Econometrics, Vol. 38 (Kluwer Academic, Boston).

Bauwens, L. and N. Hautsch (2003). Dynamic latent factor models for intensity processes. Discussion Paper 2003-103, CORE, Université Catholique de Louvain.

Berman, M. (1983). Discussion of Ogata's paper. Bulletin of the International Statistical Institute, 50, 412-422.

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 (Springer-Verlag, New York).

Brémaud, P. and L. Massoulié (1996). Stability of nonlinear Hawkes processes. Annals of Probability, 24, 1563-1588.

Brown, T. C. and M. G. Nair (1988). A simple proof of the multivariate random time change theorem for point processes. Journal of Applied Probability, 25, 210-214.

Daley, D. and D. Vere-Jones (2003). An introduction to the theory of point processes. Vol. 1 (Springer-Verlag, New York).

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 (Timberlake Consultants Ltd., London).

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. (2002). New frontiers for ARCH models. Journal of Applied Econometrics, 17, 425-446.

Engle, R. F. and A. Lunde (2003). Trades and quotes: A bivariate point process. Journal of Financial Econometrics, 1, 159-188.

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.

Hall, T. and N. Hautsch (2004). A continuous time measurement of the buy-sell pressure in a limit order book market. Discussion Paper 04-07, Institute of Economics, University of Copenhagen.

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. Handbook of Statistics, Vol. 14 (North-Holland, Amsterdam) 647-692.

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.

Hong, Y. and H. Li (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Review of Financial Studies, 18, 37-84.

Karr, A. F. (1991). Point processes and their statistical inference (Dekker, New York).

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. Lecture Notes in Mathematics, Vol. 191 (Springer Verlag, Berlin) 191-195.

Norberg, R. (2002). Vandermonde in finance. Mimeo, Department of Statistics, London School of Economics.

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

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

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

O'Hara, M. (1995). Market microstructure theory (Blackwell, Oxford).

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 (2004). Modeling comovements in trading intensities to distinguish sector and stock specific news. Discussion Paper No. 69, Tilburg University and CentER.

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

Thompson, S. B. (2004). Identifying term structure volatility from the LIBOR-swap curve. Mimeo, Department of Economics, Harvard University.

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

	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.

$\boldsymbol{\theta}_1$	**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}
$l_1(\theta_1)$	12,022		132,013		$l_2(\theta_2)$
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)	$BL^{(2)}$
$BL_2^{(1)}$	12.313	(0.655)	4.7585	(0.994)	$BL_2^{(2)}$
$ED^{(1)}$	-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. The parameters of the quote intensity and the trade intensity are listed in the first and last columns respectively. 95% CIs are shown in square parentheses. The maximised log-likelihood for the bivariate model is 12,022 + 132,013 = 144,035. The reported m-tests are described in Section 4.2; BL (BL_2) denotes the Box-Ljung test based on the levels (squares) and ED the excess dispersion test; a superscript (1) denotes a test based on the quote intensity and a (2) denotes one based on the trade intensity.

θ_1	GM dataset MLEs	MC Mean	Coverage	θ_2	GM dataset MLEs	MC Mean	Coverage
$\alpha_{11}^{(1)}$	1.725 [1.2,2.4]	1.715	0.954	$\alpha_{22}^{(1)}$	10.40 [8.9,12]	10.54	0.945
$\alpha_{11}^{(2)}$	0.140	0.231	-	$\alpha_{22}^{(2)}$	1.542 [0.96,2.5]	1.514	0.960
$\beta_{11}^{(1)}$	8.626 [5.9,13]	13.06	0.962	$\beta_{22}^{(1)}$	43.49 [32,60]	43.61	0.941
$\beta_{11}^{(2)}$	1.361 [0.87,2.1]	1.603	0.954	$\beta_{22}^{(2)}$	3.359 [2.3,5.0]	3.350	0.956
$\alpha_{12}^{(1)}$	41.02 [1.2,2.5]	41.01	0.937	$\alpha_{21}^{(1)}$	216.8 [180,261]	216.76	0.949
$\alpha_{12}^{(2)}$	4.476 [1.2,2.5]	4.589	0.925	$\alpha_{21}^{(2)}$	38.88 [27,55]	38.559	0.950
$\beta_{12}^{(1)}$	811.2 [1.2,2.5]	822.1	0.946	$\beta_{21}^{(1)}$	1552 [1156,2084]	1594	0.946
$\beta_{12}^{(2)}$	96.62 [1.2,2.5]	97.67	0.928	$\beta_{21}^{(2)}$	101.6 [70,148]	106.78	0.940
γ_{11}	24.70 [20,30]	24.73	0.957	γ_{21}	91.13 [80,103]	91.33	0.948
γ_{12}	2.893 [1.2,7.3]	3.013	0.943	γ_{22}	53.23 [44,65]	54.69	0.928
γ_{13}	0.000	0.008	-	γ_{23}	17.67 [10,31]	18.54	0.921
γ_{14}	0.000	0.030	-	γ_{24}	22.05 [16,30]	22.47	0.946
γ_{15}	0.000	0.000	-	γ_{25}	18.18 [13,25]	18.64	0.930
γ_{16}	0.000	0.000	-	γ_{26}	26.49 [21,33]	27.00	0.931
γ_{17}	0.775	0.709	-	γ_{27}	31.48 [26,39]	32.01	0.935
γ_{18}	0.496	0.787	-	γ_{28}	69.54 [60,80]	71.57	0.934

Table 3: Properties of MLEs and asymmetric CIs for the BV g-HawkesE(2) model with no dependence across days. The DGP uses as parameter values the MLEs for the GM dataset (see 2nd and 6th columns); MC Mean is the Monte Carlo mean of the simulated MLEs across replications; Coverage is the proportion of replications in which the true parameter lies in the 95% CI. Results based on 1000 replications.

		$BL^{(1)}$	$BL_2^{(1)}$	$ED^{(1)}$	$BL^{(2)}$	$BL_2^{(2)}$	$ED^{(2)}$	$BL^{(o)}$	$BL_2^{(o)}$	$ED^{(o)}$
GM dataset:	statistic	21.07	12.65	-0.618	46.17	4.295	1.283	17.72	25.57	-3.049
	p-value	(0.135)	(0.630)	(0.536)	(0.000)	(0.997)	(0.200)	(0.277)	(0.043)	(0.002)
Using θ^*	0.05	0.054	0.071	0.054	0.046	0.056	0.062	0.044	0.060	0.049
	0.01	0.014	0.022	0.013	0.005	0.008	0.022	0.005	0.017	0.008
Using $\hat{\theta}$	0.05	0.050	0.066	0.000	0.029	0.039	0.001	0.036	0.048	0.017
	0.01	0.011	0.023	0.000	0.002	0.004	0.000	0.008	0.017	0.001

Table 4: Size properties of specification tests for the BV g-HawkesE(2) model with no dependence across days. DGP as in Table 3; the m-tests of specification are as in Table 2 and the omnibus tests are also shown (denoted by a superscript o); θ^* is the true parameter vector and $\hat{\theta}$ is the MLE; rejection frequencies are shown at the 5% and 1% nominal levels. Results based on 1000 replications.

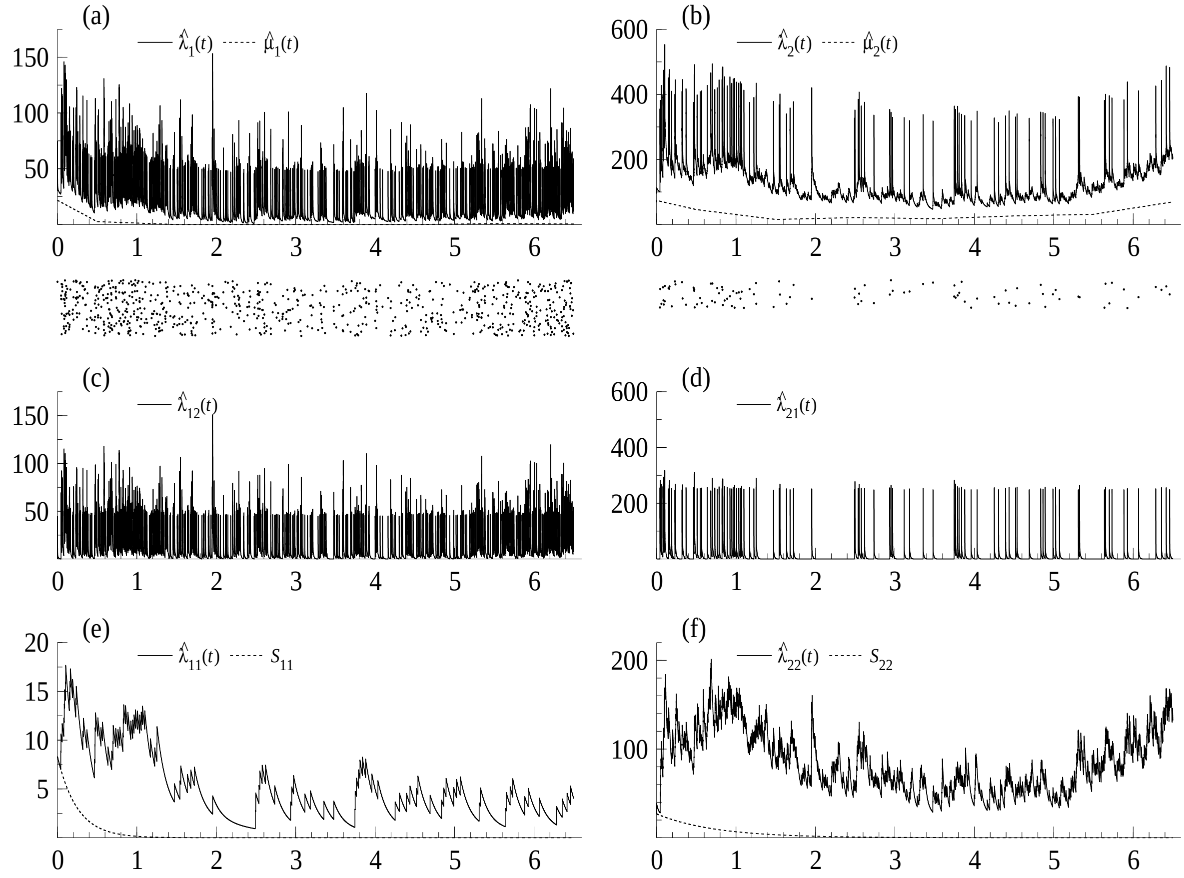

Figure 1: Components of the estimated mid-quote event and trade intensities for General Motors on 2 August 2000. The first column of panels is for the mid-quote event intensity and the second for the trade intensity. (a) the estimated total intensity, $\hat{\lambda}_1(t)$, and estimated deterministic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

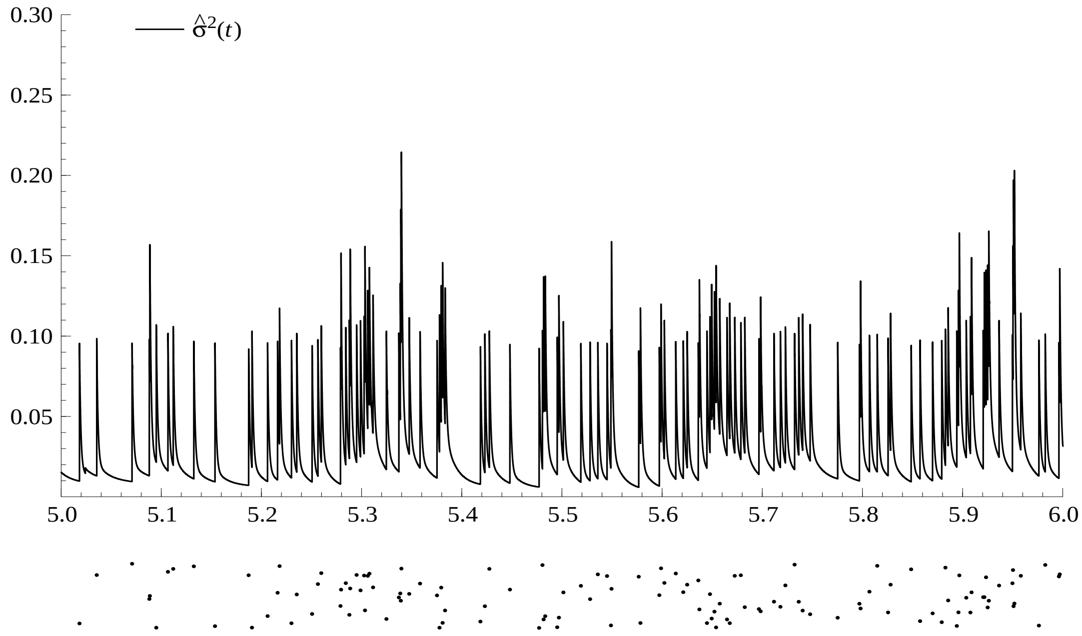

Figure 2: Estimated instantaneous volatility, $\hat{\sigma}^2(t)$, for General Motors on 2 August 2000 between 14:30 and 15:30 EST. $\hat{\sigma}^2(t)$ is based on equation (29). Also shown are the (vertically jittered) occurrence times of the trades. Volatility is measured per trading year (where one year is 252×6.5 trading hours). The horizontal axis is time measured in hours since the start of the trading day.