



Modelling financial high frequency data using point processes

Hautsch, Nikolaus; Bauwens, Luc

Publication date:
2006

Document version
Early version, also known as pre-print

Citation for published version (APA):
Hautsch, N., & Bauwens, L. (2006). *Modelling financial high frequency data using point processes*. Université catholique de Louvain.

CORE DISCUSSION PAPER

2006/80

**MODELLING FINANCIAL HIGH FREQUENCY DATA USING POINT
PROCESSES**

Luc Bauwens¹ and Nikolaus Hautsch²

September 18, 2006

Abstract

In this chapter written for a forthcoming *Handbook of Financial Time Series* to be published by Springer-Verlag, we review the econometric literature on dynamic duration and intensity processes applied to high frequency financial data, which was boosted by the work of Engle and Russell (1997) on autoregressive duration models.

Keywords: Duration, intensity, point process, high frequency data, ACD models.

JEL Classification: C41, C32.

¹CORE and Department of Economics, Université catholique de Louvain (bauwens@core.ucl.ac.be).

²University of Copenhagen, Studiestraede 6 DK-1455 Copenhagen K Denmark (Nikolaus.Hautsch@econ.ku.dk).

Work supported in part by the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, MICFINMA.

This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

Since the seminal papers by Hasbrouck (1991) and Engle and Russell (1998) the modelling of financial data at the transaction level is an ongoing topic in the area of financial econometrics. This has created a new body of literature which is often referred to as "the econometrics of (ultra-)high-frequency finance" or "high-frequency econometrics". The consideration of the peculiar properties of financial transaction data, such as the irregular spacing in time, the discreteness of price changes, the bid-ask bounce as well as the presence of serial dependence, provoked the surge of new econometric approaches. One important string of the literature deals with the irregular spacing of data in time. Taking into account the latter is indispensable whenever the full amount of information in financial transaction data has to be exploited and no loss of information due to fixed-interval aggregation schemes can be accepted. Moreover, it has been realized that the timing of trading events, such as the arrival of particular orders and trades, and the frequency in which the latter occur have information value for the state of the market and play an important role in market microstructure analysis, for the modelling of intraday volatility as well as the measurement of liquidity and implied liquidity risks.

Taking into account the irregular occurrence of transaction data requires to consider it as a point process, a so-called *financial point process*. Depending on the type of the financial "event" under consideration, we can distinguish between different types of financial point processes or processes of so-called *financial durations*. The most common types are *trade durations* and *quote durations* as defined by the time between two consecutive trade or quote arrivals, respectively. *Price durations* correspond to the time between absolute cumulative price changes of given size and can be used as an alternative volatility measure. Similarly, a *volume duration* is defined as the time until a cumulative order volume of given size is traded and captures an important dimension of market liquidity. For more details and illustrations, see Bauwens and Giot (2001) or Hautsch (2004).

One important property of transaction data is that market events are clustered over time implying that financial durations follow positively auto-correlated processes with a strong persistence. Actually, it turns out that the dynamic properties of financial durations are quite similar to those of daily volatilities. Taking into account these properties leads to different types of dynamic models on the basis of a duration representation, an intensity representation or a counting representation of a point process.

In this chapter, we review duration-based and intensity-based models of financial point processes. In Section 2, we introduce the fundamental concepts of point process theory and discuss major statistical tools. In Section 3, we review the class of *dynamic duration models*. Specifying a (dynamic) duration model is presumably the most intuitive way to characterize a point process in discrete time and has been suggested by Engle and Russell (1998), which was the starting point for a huge body of literature. Nevertheless, Russell (1999)

realized that a continuous-time setting on the basis of the intensity function constitutes a more flexible framework which is particularly powerful for the modelling of multivariate processes. Different types of *dynamic intensity models* are presented in Section 4.

2 Fundamental Concepts of Point Process Theory

In this section, we discuss important concepts and relations in point process theory which are needed throughout this chapter. In Section 2.1, we introduce the notation and basic definitions. The fundamental concepts of intensity functions, compensators and hazard rates are defined in Section 2.2, whereas in Section 2.3 different classes and representations of point processes are discussed. Finally, in Section 2.4, we present the random time change theorem which yields a powerful result for the construction of diagnostics for point process models. Most concepts discussed in this section are based upon Karr (1991).

2.1 Notation and Definitions

Let $\{t_i\}_{i \in \{1, \dots, n\}}$ denote a random sequence of increasing event times ($0 < t_1 < \dots < t_n$) associated with an orderly (simple) point process. Then, $N(t) := \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}}$ defines the right-continuous (càdlàg) counting function. Throughout this chapter, we consider only point processes which are integrable, i.e. $E[N(t)] < \infty \forall t \geq 0$. Furthermore, $\{W_i\}_{i \in \{1, \dots, n\}}$ denotes a sequence of $\{1, \dots, K\}$ -valued random variables representing K different types of events. Then, we call the process $\{t_i, W_i\}_{i \in \{1, \dots, n\}}$ an K -variate *marked* point process on $(0, \infty)$ as represented by the K sequences of event-specific arrival times $\{t_i^k\}_{i \in \{1, \dots, n^k\}}$, $k = 1, \dots, K$, with counting functions $N^k(t) := \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}} \mathbb{1}_{\{W_i = k\}}$.

The internal history of an K -dimensional point process $N(t)$ is given by the filtration \mathcal{F}_t^N with $\mathcal{F}_t^N = \sigma(N^k(s) : 0 \leq s \leq t, k \in \Xi)$, $N^k(s) = \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq s\}} \mathbb{1}_{\{W_i \in \Xi\}}$, where Ξ denotes the σ -field of all subsets of $\{1, \dots, K\}$. More general filtrations, including e.g. also processes of explanatory variables (covariates) $\{z_i\}_{i \in \{1, \dots, n\}}$ are denoted by \mathcal{F}_t with $\mathcal{F}_t^N \subseteq \mathcal{F}_t$.

Define $x_i := t_i - t_{i-1}$ with $i = 1, \dots, n$ and $t_0 := 0$ as the inter-event duration from t_{i-1} until t_i . Furthermore, $x(t)$ with $x(t) := t - t_{\check{N}(t)}$, with $\check{N}(t) := \sum_{i \geq 1} \mathbb{1}_{\{t_i < t\}}$ denoting the left-continuous counting function, is called the backward recurrence time. It is a left-continuous function that grows linearly through time with discrete jumps back to zero *after* each arrival time t_i . Finally, let $\theta \in \Theta$ denote model parameters.

2.2 Compensators, Intensities, and Hazard Rates

In martingale-based point process theory, the concept of *compensators* plays an important role. Using the property that an \mathcal{F}_t -adapted point process $N(t)$

is a submartingale¹, it can be decomposed into a zero mean martingale $M(t)$ and a (unique) \mathcal{F}_t -predictable increasing process, $\tilde{A}(t)$, which is called the compensator of $N(t)$ and can be interpreted as the local conditional mean of $N(t)$ given the past. This decomposition is typically referred to the Doob-Meyer decomposition, see also Theorem 2.14 in Karr (1991).

Define $\lambda(t)$ as a scalar, positive \mathcal{F}_t -predictable process, i.e. $\lambda(t)$ is adapted to \mathcal{F}_t , and left-continuous with right hand limits. Then, $\lambda(t)$ is called the (\mathcal{F}_t -conditional) *intensity* of $N(t)$ if

$$\tilde{A}(t) = \int_0^t \lambda(u) du, \quad (1)$$

where $\tilde{A}(t)$ is the (unique) compensator of $N(t)$. This relationship emerges from the interpretation of the compensator as integrated (conditional) hazard function as established by Theorem 2.18 and Definition 2.28 in Karr (1991). Consequently, $\lambda(t)$ can be also defined by the relation

$$\mathbb{E}[N(s) - N(t) | \mathcal{F}_t] = \mathbb{E} \left[\int_t^s \lambda(u) du \middle| \mathcal{F}_t \right] \quad (2)$$

which has to hold (almost surely) for all t, s with $0 \leq t \leq s$. Letting $s \downarrow t$ leads to the heuristic representation which is more familiar in classical duration analysis. Then, $\lambda(t)$ is obtained by

$$\lambda(t+) := \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathbb{E}[N(t + \Delta) - N(t) | \mathcal{F}_t], \quad (3)$$

where $\lambda(t+) := \lim_{\Delta \downarrow 0} \lambda(t + \Delta)$. In case of a stationary point process, $\bar{\lambda} := \mathbb{E}[dN(t)]/dt = \mathbb{E}[\lambda(t)]$ is constant.

Equation (3) manifests the close analogy between the intensity function and the *hazard function* which is given by

$$h(x) := f(x)/S(x) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \Pr[x \leq X < x + \Delta | X \geq x] \quad (4)$$

with x denoting the (inter-event) duration as represented by the realization of a random variable X with probability density function $f(x)$, survivor function $S(x) = 1 - F(x)$, and cumulative distribution function (cdf) $F(x) = \Pr[X \leq x]$. Whereas the intensity function is defined in (continuous) calendar time, the hazard rate is typically defined in terms of the length of a duration x and is a key concept in (cross-section) survival analysis. For more details regarding the analogy between intensities and hazard rates, see Theorems 2.30 and 2.31 in Karr (1991).

¹ An \mathcal{F}_t -adapted càdlàg process $N(t)$ is a submartingale if $\mathbb{E}[|N(t)|] < \infty$ for each t and if $s < t$ implies that $\mathbb{E}[N(t) | \mathcal{F}_s] \geq N(s)$. See e.g. Karr (1991).

2.3 Types and Representations of Point Processes

The simplest type of point process is the homogeneous Poisson process defined by

$$\Pr[(N(t + \Delta) - N(t)) = 1 | \mathcal{F}_t] = \lambda\Delta + o(\Delta), \quad (5)$$

$$\Pr[(N(t + \Delta) - N(t)) > 1 | \mathcal{F}_t] = o(\Delta), \quad (6)$$

with $\Delta \downarrow 0$. Then, $\lambda > 0$ is called the Poisson rate corresponding to the (constant) intensity. Accordingly, equations (5) and (6) define the *intensity representation* of a Poisson process. A well-known property of homogenous Poisson processes is that the inter-event waiting times $x_i = t_i - t_{i-1}$ are independently exponentially distributed, leading to the *duration representation*. In this context, λ is the hazard rate of the exponential distribution. Furthermore, it can be shown (see e.g. Lancaster (1997)) that the number of events in an interval $(a, b]$, $N(a, b) := N(b) - N(a)$ is Poisson distributed with $\Pr[N(a, b) = k] = \exp[-\lambda(b - a)][\lambda(b - a)]^k/k!$, yielding the *counting representation*. All three representations of a Poisson process can be used as the starting point for the specification of a point process model.

Throughout this chapter we associate the term *duration models* to a model of the (discrete-time) duration process observable at the event-times $\{t_i\}_{i=1, \dots, n}$. Then, researchers parameterize the conditional distribution function $F(x_i | \mathcal{F}_{t_{i-1}})$ or, alternatively, the conditional hazard rate $h(x_i | \mathcal{F}_{t_{i-1}})$. Generally, such a model should aim, in particular, at fitting the dynamical and distributional properties of durations. The latter is often characterized by the excess dispersion, corresponding to the ratio between the standard deviation to the mean. In classical hazard rate models employed in traditional survival analysis, the hazard rate is typically parameterized in terms of covariates, see e.g. Kalbfleisch and Prentice (1980), Kiefer (1988) or Lancaster (1997). The most well-known hazard model is the *proportional hazard* model introduced by Cox (1972) and is given by

$$h(x|z; \theta) = h_0(x|\gamma_1)g(z, \gamma_2), \quad (7)$$

where $\theta = (\gamma_1, \gamma_2)$, $h_0(\cdot)$ denotes the so-called baseline hazard rate and $g(\cdot)$ is a function of the covariates z and parameters γ_2 . The baseline hazard rate may be parameterized in accordance with a certain distribution, like e.g., a Weibull distribution with parameters $\lambda, p > 0$ implying

$$h_0(x|\gamma_1) = \lambda p (\lambda x)^{p-1}. \quad (8)$$

For $p = 1$ we yield the exponential case $h_0(x|\gamma_1) = \lambda$, implying a constant hazard rate. Alternatively, if $p > 1$, $\partial h_0(x|\gamma_1)/\partial x > 0$, i.e. the hazard rate is increasing with the length of the spell which is referred to as "positive duration dependence". In contrast, $p < 1$ implies negative duration dependence. Non-monotonic hazard rates can be obtained with more flexible distributions,

like the generalized F and particular cases thereof, including the generalized gamma, Burr, Weibull and log-logistic distributions. We refer to the Appendix to Chapter 3 of Bauwens and Giot (2001) and to the Appendix of Hautsch (2004) for definitions and properties. Alternatively, the baseline hazard may be left unspecified and can be estimated nonparametrically, see Cox (1975).

An alternative type of duration model is the class of *accelerated failure time* (AFT) models given by

$$h(x|z; \theta) = h_0[xg(z, \gamma_2)|\gamma_1]g(z, \gamma_2). \quad (9)$$

Here, the effect of the exogenous variables is to accelerate or to decelerate the time scale on which the baseline hazard h_0 is defined. As illustrated in Section 3.1, AFT-type models are particularly attractive to allow for autocorrelated duration processes.

Because of their discrete-time nature, duration models cannot be used whenever the information set has to be updated *within* a duration spell, e.g. caused by time-varying covariates or event arrivals in other point processes. For this reason, (discrete-time) duration models are typically used in an univariate framework.

Whenever a continuous-time modelling is advantageous (as e.g. to account for the asynchronous event arrivals in a multivariate framework), it is more natural to specify the intensity function directly. This class of models is referred to as *intensity models*. One important extension of a homogenous Poisson process is to allow the intensity to be directed by a real-valued, non-negative (stationary) random process $\lambda^*(t)$ with (internal) history \mathcal{F}_t^* leading to the class of *doubly stochastic Poisson processes* (or Cox process). In particular, $N(t)$ is called a Cox process directed by $\lambda^*(t)$ if conditional on $\lambda^*(t)$, $N(t)$ is a Poisson process with mean $\lambda^*(t)$, i.e. $\Pr[N(a, b) = k | \mathcal{F}_t^*] = \exp[-\lambda^*(t)] [\lambda^*(t)]^k / k!$. The doubly stochastic Poisson process yields a powerful class of probabilistic models with applications in seismology, biology and economics. For instance, specifying $\lambda^*(t)$ in terms of an autoregressive process yields a dynamic intensity model which is particularly useful to capture the clustering in financial point processes. For a special type of doubly stochastic Poisson process see Section 4.2.

A different generalization of the Poisson process is obtained by specifying $\lambda(t)$ as a (linear) self-exciting process given by

$$\lambda(t) = \omega + \int_0^t w(t-u) dN(u) = \omega + \sum_{t_i < t} w(t-t_i), \quad (10)$$

where ω is a constant, $w(s)$ denotes a non-negative weight function, and $\int_0^t w(s) dN(s)$ is the stochastic Stieltjes integral of the process w with respect to the counting process $N(t)$. The process (10) was proposed by Hawkes (1971) and is therefore named a *Hawkes process*. If $w(s)$ declines with s , then, the process is *self-exciting* in the sense that $\text{Cov}[N(a, b), N(b, c)] > 0$, where

$0 < a \leq b < c$. Different types of Hawkes processes and their applications to financial point processes are presented in Section 4.1. A further type of intensity models which is relevant in the literature of financial point processes is given by a specification where the intensity itself is driven by an autoregressive process which is updated at each point of the process. This leads to a special type of point process models which does not originate from the classical point process literature but originates from the ACD literature reviewed in Section 2 and brings time series analysis into play. Such a process is called an *autoregressive conditional intensity* model and is considered in Section 4.2.

Finally, starting from the counting representation of a Poisson process leads to the class of *count data models*. Dynamic extensions of Poisson processes in terms of counting representations are not surveyed in this chapter. Some references reflecting the diversity of approaches are Rydberg and Shephard (2003), Heinen and Rengifo (2003), Liesenfeld et al. (2006), and Quoreshi (2006).

2.4 The Random Time Change Theorem

One fundamental result of martingale-based point process theory is the (multivariate) random time change theorem by Meyer (1971) which allows to transform a wide class of point processes to a homogeneous Poisson process:

Theorem (Meyer, 1971, Brown and Nair, 1988): *Assume a multivariate point process $(N^1(t), \dots, N^K(t))$ is formed from the event times $\{t_i^k\}_{i \in \{1, \dots, n^k\}}$, $k = 1, \dots, K$, and has continuous compensators $(\tilde{\Lambda}(t)^1, \dots, \tilde{\Lambda}(t)^K)$ with $\tilde{\Lambda}^k(\infty) = \infty$ for each $k = 1, \dots, K$, then the point processes formed from $\{A^k(t_i^k)\}_{i \in \{1, \dots, n^k\}}$, $k = 1, \dots, K$, are independent Poisson processes with unit intensity.*

Proof: See Meyer (1971) or Brown and Nair (1988) for a more accessible and elegant proof.

Define $\tau^k(t)$ as the (\mathcal{F}_t) -stopping time obtained by the solution of $\int_0^{\tau^k(t)} \lambda^k(s) ds = t$. Applying the random time change theorem to (1) implies that the point processes $\tilde{N}^k(t)$ with $\tilde{N}^k(t) := N^k(\tau^k(t))$ are independent Poisson processes with unit intensity and event times $\{\tilde{A}^k(t_i^k)\}_{i \in \{1, \dots, n^k\}}$ for $k = 1, \dots, K$. Then, the so-called integrated intensities

$$A^k(t_{i-1}^k, t_i^k) := \int_{t_{i-1}^k}^{t_i^k} \lambda^k(s) ds = \tilde{\Lambda}^k(t_i^k) - \tilde{\Lambda}^k(t_{i-1}^k) \quad (11)$$

correspond to the increments of independent Poisson processes for $k = 1, \dots, K$. Consequently, they are independently standard exponentially distributed across i and k . For more details, see also Bowsler (2002). The random time change theorem plays an important role in order to construct diagnostic tests for point process models (see Section 4.3) or to simulate point processes (see e.g. Giesecke and Tomecek (2005)).

3 Dynamic Duration Models

In this section, we discuss univariate dynamic models for the durations between consecutive (financial) events. In Section 3.1, we review in detail the class of autoregressive conditional duration (ACD) models, which is by far the most used class in the literature on financial point processes. In Section 3.2, we briefly discuss statistical inference for ACD models. In Section 3.3, we present other dynamic duration models, where in the last section we review some applications.

3.1 ACD Models

The class of ACD models has been introduced by Engle and Russell (1997, 1998) and Engle (2000). In order to keep the notation simple, define x_i in the following as the inter-event duration which is standardized by a seasonality function $s(t_i)$, i.e. $x_i := (t_i - t_{i-1})/s(t_i)$. The function $s(t_i)$ is typically parameterized according to a spline function capturing time-of-day or day-of-week effects. Time-of-day effects arise because of systematic changes of the market activity throughout the day and due to opening of other related markets. In most approaches $s(t_i)$ is specified according to a linear or cubic spline function and is estimated separately in a first step yielding seasonality adjusted durations x_i . Alternatively, a non-parametric approach has been proposed by Veredas et al. (2002). For more details and examples regarding seasonality effects in financial duration processes, we refer the reader to Chapter 2 of Bauwens and Giot (2001) or to Chapter 3 of Hautsch (2004).

The key idea of the ACD model is to model the (seasonally adjusted) durations $\{x_i\}_{i=1,\dots,n}$ in terms of a multiplicative error term model in the spirit of Engle (2002), i.e.

$$x_i = \Psi_i \epsilon_i, \quad (12)$$

where Ψ_i denotes a function of the past durations (and possible covariates), and ϵ_i defines an i.i.d. random variable for which it is assumed that

$$E(\epsilon_i) = 1, \quad (13)$$

so that Ψ_i corresponds to the conditional duration mean (the so-called "conditional duration") with $\Psi_i := E(x_i | \mathcal{F}_{t_{i-1}})$. The ACD model can be rewritten in terms of the intensity function as

$$\lambda(t | \mathcal{F}_t) = \lambda_\epsilon \left(\frac{x(t)}{\Psi_{\tilde{N}(t)+1}} \right) \frac{1}{\Psi_{\tilde{N}(t)+1}}, \quad (14)$$

where $\lambda_\epsilon(s)$ denotes the hazard function of the ACD error term. This formulation clearly demonstrates that the ACD model belongs to the class of AFT models. Assuming ϵ_i to be standard exponentially distributed yields the

so-called *Exponential ACD* model. More flexible specifications arise by assuming ϵ_i to follow a more general distribution, see the discussion after eqn (8). It is evident that the ACD model is the counter-part to the GARCH model (Bollerslev (1986)) for duration processes. Not surprisingly, many results and specifications from the GARCH literature have been carried over to the ACD literature.

The conditional duration, Ψ_i , is defined as a function Ψ of the information set $\mathcal{F}_{t_{i-1}}$ and provides therefore the vehicle for incorporating the dynamics of the duration process. In this respect it is convenient to use an ARMA-type structure of order (p, q) , whereby

$$\Psi_i = \Psi(\Psi_{i-1}, \dots, \Psi_{i-p}, x_{i-1}, \dots, x_{i-q}). \quad (15)$$

For simplicity, we limit the exposition in the sequel to the case $p = q = 1$.

First Generation ACD Models

We group under this label the linear ACD model of Engle and Russell (1998), the logarithmic ACD models of Bauwens and Giot (2000), the augmented ACD classes of Fernandes and Grammig (2006) and Hautsch (2006), and the semiparametric ACD model suggested by Hautsch (2006). All these models have their GARCH counterparts. Analytical results are available for some of these models about stationarity conditions, conditions for existence of moments, and formulas of the moments.

The first model put forward in the literature is the *linear ACD model*, which specializes (15) as

$$\Psi_i = \omega + \beta\Psi_{i-1} + \alpha x_{i-1}. \quad (16)$$

Since Ψ_i must be positive, the restrictions $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$ are usually imposed. It is also assumed that $\beta = 0$ if $\alpha = 0$, otherwise β is a redundant parameter. The process defined by (12), (13) and (16) is known to be covariance-stationary if

$$(\alpha + \beta)^2 - \alpha^2\sigma^2 < 1, \quad (17)$$

where $\sigma^2 = \text{Var}(\epsilon_i) < \infty$, and to have the following moments and autocorrelations:

- (1) $E(x_i) = \mu_x = \omega / (1 - \alpha - \beta)$,
- (2) $\text{Var}(x_i) = \sigma_x^2 = \mu_x^2 \sigma^2 \frac{1 - \beta^2 - 2\alpha\beta}{1 - (\alpha + \beta)^2 - \alpha^2\sigma^2}$,
- (3) $\rho_1 = \frac{\alpha(1 - \beta^2 - \alpha\beta)}{1 - \beta^2 - 2\alpha\beta}$ and $\rho_n = (\alpha + \beta)\rho_{n-1}$ ($n \geq 2$).

The condition (17) ensures the existence of the variance. These results are akin to those for the GARCH(1,1) zero-mean process. They can be generalized to ACD(p, q) processes when $p, q > 1$. It is usually found empirically that the

estimates of the parameters are such that $\alpha + \beta$ is in the interval (0.85,1) while α is in the interval (0.01,0.15). Since the ACD(1,1) model can be written as

$$x_i = \omega + (\alpha + \beta)x_{i-1} + u_i - \beta u_{i-1}, \quad (18)$$

where $u_i = x_i - \Psi_i$ is a martingale difference innovation, the resulting autocorrelation function (ACF) is that of an ARMA(1,1) process that has AR and MA roots close to each other. This type of parameter configuration generates the typical ACF shape of clustered data. Nevertheless, the ACF decreases at a geometric rate, though it is not uncommon to find duration series with an ACF that decreases at a hyperbolic rate. This tends to happen for long series and may be due to parameter changes that give the illusion of long memory in the process. In order to allow for long range dependence in financial duration processes, the ACD model has been extended to a *fractionally integrated ACD* model by Jasiak (1998) and has been applied to IBM and Alcatel trade durations.

A drawback of the linear ACD model is that it is difficult to allow Ψ_i to depend on functions of covariates without violating the non-negativity restriction. For this reason, Bauwens and Giot (2000) propose a class of *logarithmic ACD models*, where no parametric restrictions are needed to ensure positiveness of the process:

$$\ln \Psi_i = \omega + \beta \ln \Psi_{i-1} + \alpha g(\epsilon_{i-1}), \quad (19)$$

where $g(\epsilon_{i-1})$ is either $\ln \epsilon_{i-1}$ (log-ACD of type I) or ϵ_{i-1} (type II). Using this setting, it is convenient to augment Ψ_i by functions of covariates, see e.g. Bauwens and Giot (2001). The stochastic process defined by (12), (13) and (19) is covariance-stationary if

$$\beta < 1, \quad \mathbb{E}\{\epsilon_i \exp[\alpha g(\epsilon_i)]\}, \quad \mathbb{E}\{\exp[2\alpha g(\epsilon_i)]\} < \infty, \quad (20)$$

and has the following moments and autocorrelations, see Fernandes and Grammig (2006) or Section 3.2 in Bauwens and Giot (2001):

- (1) $\mathbb{E}(x_i) = \mu_x = \exp\left(\frac{\omega}{1-\beta}\right) \prod_{j=1}^{\infty} \mathbb{E} \exp[\alpha \beta^{j-1} g(\epsilon_i)],$
- (2) $\text{Var}(x_i) = \sigma_x^2 = (1 + \sigma^2) \exp\left(\frac{2\omega}{1-\beta}\right) \prod_{j=1}^{\infty} \mathbb{E} \exp[2\alpha \beta^{j-1} g(\epsilon_i)] - \mu_x^2,$
- (3) For $n \geq 1,$

$$\rho_n = \frac{\mathbb{E}[\epsilon_i e^{\alpha \beta^{n-1} g(\epsilon_i)}] \prod_{j=1}^{n-1} \mathbb{E}[e^{\alpha \beta^{j-1} g(\epsilon_i)}] \prod_{j=1}^{\infty} \mathbb{E}[e^{\alpha(1+\beta^n)\beta^{j-1} g(\epsilon_i)}]}{(1 + \sigma^2) \prod_{j=1}^{\infty} \mathbb{E} \exp[2\alpha \beta^{j-1} g(\epsilon_i)] - [\mathbb{E}(\epsilon_i)]^2 \prod_{j=1}^{\infty} [\mathbb{E} e^{\alpha \beta^{j-1} g(\epsilon_i)}]^2} - \frac{\mu_x^2}{\sigma_x^2}.$$

An even more flexible type of ACD model is the *augmented ACD* (AACD) model introduced by Fernandes and Grammig (2006). Here, Ψ_i is specified in terms of a Box-Cox transformation yielding

$$\Psi_i^{\delta_1} = \omega + \beta \Psi_{i-1}^{\delta_1} + \alpha \Psi_{i-1}^{\delta_1} [|\epsilon_{i-1} - \xi| - \rho(\epsilon_{i-1} - \xi)]^{\delta_2}, \quad (21)$$

where $\delta_1 > 0$, $\delta_2 > 0$, ξ , and ρ are parameters. The so-called news impact function $[|\epsilon_{i-1} - \xi| - \rho(\epsilon_{i-1} - \xi)]^{\delta_2}$ allows for a wide variety of shapes of the curve tracing the impact of ϵ_{i-1} on Ψ_i for a given value of Ψ_{i-1} and the remaining parameters. The parameter ξ is a shift parameter and the parameter ρ is a rotation parameter. If $\xi = 0$, the impact curve is increasing (concave if $\delta_2 \leq 1$, convex if $\delta_2 \geq 1$). If ξ and ρ are both positive, the impact can decrease if $\epsilon_{i-1} < \xi$ and increase if $\epsilon_{i-1} > \xi$. If $\xi = \rho = 0$, the linear ACD model is obtained by setting $\delta_1 = \delta_2 = 1$, the type I logarithmic ACD model by letting δ_1 and δ_2 tend to 0, and the type II version by letting δ_1 tend to 0 and setting $\delta_2 = 1$. Using results of Carrasco and Chen (2002), Fernandes and Grammig (2006) provide sufficient conditions for the strict stationarity and β -mixing with exponential decay of the process defined by (12), (13) and (21). These conditions include $|\beta| < 1$ and some moment conditions. There is no analytical formula for moments of x_i in the case of the general AACD equation, although some recursion formulas are available for moments of $x_i^{\delta_1}$ (but not for x_i itself). As pointed above, analytical formulas are available for particular cases like the linear and logarithmic versions. Fernandes and Grammig (2006) compare different versions of the AACD model using IBM price durations arising from trading at the New York Stock Exchange. Their main finding is that "letting δ_1 free to vary and accounting for asymmetric effects (by letting ξ and ρ free) seem to operate as substitute sources of flexibility". However, empirical studies with other data series are needed to validate this conclusion.

While the basic ACD model, the type II Log-ACD model, and the AACD model is based on multiplicative stochastic components (in the sense that the process is updated by some multiplicative function of ϵ_i and Ψ_i), Hautsch (2006) proposes an even more general class of ACD models including multiplicative stochastic components as well as additive stochastic components. It nests a wide range of special cases including the specifications outlined above and is given by replacing the news impact function in (21) by

$$\alpha \Psi_{i-1}^{\delta_1} [|\epsilon_{i-1} - \xi| - \rho(\epsilon_{i-1} - \xi)]^{\delta_2} + \nu [|\epsilon_{i-1} - \xi| - \rho(\epsilon_{i-1} - \xi)],$$

where ν is an additional parameter capturing the impact of the additive stochastic component. Clearly, ν is only defined if $\delta_1 \neq 0$. For $\nu = 0$, we obtain the AACD model. For $\alpha = 0$, $\xi = 0$ and $\rho = 0$, it nests the Box-Cox ACD model introduced by Hautsch (2003). For $\alpha = 0$, $\delta_1 \rightarrow 0$, and $\rho = 1$, it corresponds to the so-called EXponential ACD model proposed by Dufour and Engle (2000) implying a kinked news impact function.

Finally, Hautsch (2006) suggests a semiparametric ACD model corresponding to the counterpart to the semiparametric GARCH model proposed by Engle and Ng (1993). Here, the news impact function is specified in terms of a linear spline function based on the support of ϵ_i . Hautsch (2006) illustrates that the high flexibility of the latter two models is needed in order to appropriately capture the dynamic properties of financial durations. Since both types of ACD models also belong to the class of models defined by Carrasco and

Chen (2002) their results regarding stationarity and unconditional moments can be applied in this framework as well.

Second Generation ACD Models

The class of second generation ACD models consists of regime-shifting ACD models and mixture ACD models. The mixture can be only on the error distribution, but it can also involve the dynamic component.

Zhang et al. (2001) propose a *threshold ACD model* (TACD), wherein the ACD equation and the error distribution change according to a threshold variable such as the previous duration. For J regimes indexed by $j = 1, \dots, J$, the model is defined as

$$x_i = \Psi_i \epsilon_i^{(j)}, \quad (22)$$

$$\Psi_i = \omega_j + \beta_j \Psi_{i-1} + \alpha_j x_{i-1} \quad (23)$$

when $x_{i-1} \in [r_{j-1}, r_j)$, and $0 = r_0 < r_1 < \dots < r_J = \infty$ are the threshold parameters. The superscript (j) on the error ϵ_i indicates that the distribution or its parameters can vary with the regime operating at observation i . This model can be viewed as a mixture of J ACD models, where the probability to be in regime j at i is equal to 1 and the probabilities to be in each of the other regimes is equal to 0. The estimation of TACD models for given values of the threshold parameters is easily performed by ML. This is combined with a grid search over the threshold parameters to obtain the ML estimates of all the parameters.

While the TACD model implies discrete transitions between the individual regimes, Meitz and Teräsvirta (2006) propose a class of *smooth transition ACD (STACD) models* which are not encompassed by the augmented ACD class but generalize the linear and logarithmic ACD models in a specific way. For example, a STACD model is specified as

$$\Psi_i = \omega + \beta \Psi_{i-1} + \alpha x_{i-1} + (\omega_* + \alpha_* x_{i-1}) G(\ln x_{i-1}) \quad (24)$$

where ω_* and α_* are additional parameters and $G(\ln x_{i-1}; \zeta, c)$ is a bounded and positive ‘transition’ function. For example, this can be the logistic function

$$G(\ln x_{i-1}; \zeta, \tau_1, \dots, \tau_K) = \left[1 + \exp \left(-\zeta \prod_{k=1}^K (\ln x_{i-1} - \tau_k) \right) \right]^{-1} \quad (25)$$

where $\zeta > 0$, $\tau_1 \leq \dots \leq \tau_K$ are parameters, and the integer K determines the flexibility of the transition function. For example, with $K = 2$, the transition function can be close to 1 when x_{i-1} is small, decrease more or less fastly to be close to 0 over a range of intermediate values, and finally increase and stay close to 1 for large durations. Notice that the logarithm of x_{i-1} is used as argument rather than x_{i-1} since the logistic function is defined on the whole

real axis. The two regime TACD model with the restriction $\epsilon_i^{(1)} = \epsilon_i^{(2)}$ is a STACD model wherein $K = 1$, $\tau_1 = \ln r_1$, and ζ tends to ∞ in the transition function (25), which then becomes the indicator function for $x_{i-1} > r_1$. The three regime TACD model can likewise be obtained as the limit of a STACD model with $K = 2$ in the transition function and a unique error distribution. Conditions for strict stationarity, ergodicity, and existence of moments for the STACD (and other ACD) models are provided in Meitz and Saikkonen (2004) using the theory of Markov chains. It is clear that a motivation for the STACD model is, like for the AACD, to allow for a nonlinear impact of the past duration on the next expected duration. The variable driving the transition function can also be replaced by a measure of time (e.g. the calendar time) and the transition can also be applied to the lagged expected duration term, yielding a *time-varying ACD model*. This requires to shift the value of the transition function (25), see Meitz and Teräsvirta (2006) for details.

Another category of dynamic mixture models is the *Markov-switching ACD* class (MS-ACD). Like in TACD models, each regime has its own ACD-type equation—see (23)—, but the transitions between regimes are governed by a Markov chain. Thus, with two regimes, there are two probabilities: π_{11} , the probability that x_i is governed by regime 1 when x_{i-1} is in regime 1, and π_{22} likewise for regime 2. These probabilities are fixed parameters in Hujer et al. (2002). More generally, they could be functions of observable predetermined variables. In a different but related *regime-switching ACD* model (RS-ACD), Bauwens et al. (2006a) specify the time-varying probability $\pi_{i,j}$ of observation x_i to be in regime j as a logistic transformation of the previous duration. They show that if the first regime is stable (in the sense that $\alpha_1 + \beta_1 < 1$), the ACD parameters are positive in all regimes, and $\pi_{i,1}$ tends to 1 when x_{i-1} tends to infinity, the process is geometrically ergodic, strictly stationary and β -mixing with exponential decay. The TACD model is obtained as a particular case of the RS-ACD model, when one probability is equal to 1 and the other probabilities are equal to 0.

Drost and Werker (2004) propose to combine one of the previous ACD equations for the conditional duration with an unspecified distribution for ϵ_i , yielding a class of *semi-parametric ACD models*. De Luca and Zuccolotto (2003) and De Luca and Gallo (2004) propose to use a fixed parameter ACD equation together with a mixture of two distributions for ϵ_i , i.e.

$$f(\epsilon_i) = \pi f_I(\epsilon_i) + (1 - \pi) f_U(\epsilon_i). \quad (26)$$

where π is a parameter in $(0, 1)$. They use in practice a mixture of exponential distributions with expected values c_I and c_U . Imposing (13) links the parameters by the relation $\pi c_I + (1 - \pi) c_U = 1$. This model can be named a *static mixture ACD model* (SM-ACD). Using a mixture of Fisher distributions, Hujer and Vuletic (2005b) report that the unconditional distribution of some trade durations is well fitted by such a model whereas it is not by an ACD model with a single Fisher distribution, see also Hujer and Vuletic (2005a). De Luca

and Gallo (2006) extend the SM-ACD idea by using a time-varying probability in the mixture, i.e. they let π vary with i through a logistic function that depends on variables known at the time point (t_{i-1}) where the duration x_i starts. Assuming furthermore that c_I varies with i and c_U allows to impose a unit mean for the mixture distribution and implies that the variance is a function of the predetermined variables. In an application of their model to IBM mid-quote price durations from the TAQ data base, De Luca and Gallo (2006) use the trading intensity (i.e. the number of trades divided by the duration over which they occur) and the average volume (i.e. the ratio of trade volume and the number of trades recorded between t_{i-1} and t_i), with one lag, as predetermined variables in the probability function and in the ACD equation.

3.2 Statistical Inference

The estimation of ACD models can be easily performed by maximum likelihood (ML). Engle (2000) demonstrates that the results by Bollerslev and Wooldridge (1992) on the quasi-maximum likelihood (QML) property of the Gaussian GARCH(1,1) model can be carried over to the Exponential-ACD(1,1) model. Then, QML estimates are obtained by maximizing the quasi log likelihood function given by

$$\ln \mathcal{L}(\theta; \{x_i\}_{i=1, \dots, n}) = - \sum_{i=1}^n \left[\ln \Psi_i + \frac{x_i}{\Psi_i} \right]. \quad (27)$$

For more details we refer to Chapter 3 of Bauwens and Giot (2001), Chapter 5 of Hautsch (2004), and to the survey of Engle and Russell (2005).

Residual diagnostics and goodness-of-fit tests are straightforwardly performed by evaluating the stochastic properties of the ACD residuals $\hat{\epsilon}_i = x_i / \hat{\Psi}_i$. The dynamic properties are easily analyzed based on Portmanteau statistics or tests against independence such as proposed by Brock et al. (1996). The distributional properties can be evaluated based on Engle and Russell's (1998) test for no excess dispersion using the asymptotically standard normal test statistic $\sqrt{n/8} \hat{\sigma}^2$, where $\hat{\sigma}^2$ denotes the empirical variance of the residual series. Dufour and Engle (2000) and Bauwens et al. (2004) evaluate the model's goodness-of-fit based on the evaluation of density forecasts using the probability integral transform as proposed by Diebold et al. (1998). A nonparametric test against distributional misspecification is proposed by Fernandes and Grammig (2005) based on the work of Aït-Sahalia (1996). Statistics that exclusively test for misspecifications of the conditional mean function Ψ_i have been worked out by Meitz and Teräsvirta (2006) using the Lagrange Multiplier principle and by Hautsch (2006) using (integrated) conditional moment tests. A common result is that too simple ACD specifications, such as the ACD or Log-ACD model are not flexible enough to adequately capture the properties of observed financial durations.

While the class of first generation ACD models can be estimated by (Q)ML, several second generation models require computationally more demanding methods. The dynamic mixture ACD models are impossible to estimate by ML due to the path dependence problem of conditional durations. A similar issue arises in MS-GARCH models, see Gray (1996). The estimation can be tackled by Bayesian inference, by enlarging the parameter space with latent discrete random variables (one per observation, each taking value j with probability $\pi_{i,j}$) and using a Gibbs sampling algorithm to simulate the parameters given the latent variables and the data, and the latent variables given the parameters and the data. We refer to Bauwens et al. (2006a) for RS-ACD models, to Bauwens et al. (2006b) for RS-GARCH models, and to Henneke et al. (2006) for MS-GARCH models. These dynamic mixture models are at the frontier of current research.

3.3 Other Models

Hamilton and Jordà (2002) propose an extension of the ACD model by allowing to account for covariates which might change during a duration spell (time-varying covariates). The key idea of their so-called autoregressive conditional hazard (ACH) model is to rely on the fact that in the ACD model with exponential error distribution the intensity corresponds to the inverse of the conditional duration, i.e. $\lambda(t) = \Psi_{\tilde{N}(t)+1}^{-1}$. They extend this expression by a function of variables which are known at time $t - 1$,

$$\lambda(t) = \frac{1}{\Psi_{\tilde{N}(t)+1} + z'_{t-1}\gamma}, \quad (28)$$

where z_t are time-varying covariates which are updated during a duration spell. In this sense, the ACH model can be seen as a combination of a duration model and an intensity model. In Hamilton and Jordà (2002), the ACH model is applied to study changes of the level of the federal funds rate target in the U.S.

An alternative intensity-based model is introduced by Gerhard and Hautsch (2005). They propose a dynamic extension of a Cox (1972) proportional intensity model, where the baseline intensity $\lambda_0(t)$ is non-specified. Their key idea is to exploit the stochastic properties of the integrated intensity and to re-formulate the model in terms of a regression model with unknown left-hand variable and Gumbel distributed error terms – see Kiefer (1988) for a nice illustration of this relation. To identify the unknown baseline intensity at discrete points, Gerhard and Hautsch follow the idea of Han and Hausman (1990) and formulate the model in terms of an ordered response model based on categorized durations. In order to allow for serial dependence in the duration process, the model is extended by an observation-driven ARMA dynamic based on generalized errors. As a result, the resulting semiparametric autoregressive conditional proportional intensity (ACPI) model allows to capture

serial dependence in duration processes and to estimate conditional failure probabilities without requiring explicit distributional assumptions.

Bauwens and Veredas (2004) propose the stochastic conditional duration model (SCD) as an alternative to ACD-type models. The SCD model relates to the logarithmic ACD model in the same way as the stochastic volatility model relates to the exponential GARCH model of Nelson (1991). Thus the model is defined by equations (12), (13), and

$$\ln \Psi_i = \omega + \beta \ln \Psi_{i-1} + \gamma \epsilon_{i-1} + u_i, \quad (29)$$

where u_i is belongs to an iid process and is $N(0, \sigma^2)$ distributed. The processes $\{u_i\}$ is assumed to be independent of the process $\{\epsilon_i\}$. The set of possible distributions for the duration innovations ϵ_i is the same as for ACD models. This model allows for a rich class of hazard functions for x_i through the interplay of two distributions. The latent variable Ψ_i may be interpreted as being inversely related to the information arrival process which triggers bursts of activity on financial markets. The ‘leverage’ term $\gamma \epsilon_{i-1}$ in (29) is added by Feng et al. (2004) to allow for an intertemporal correlation between the observable duration and the conditional duration, and the correlation is found to be positive. Bauwens and Veredas (2004) use a logarithmic transformation of (12) and employ QML estimation with the Kalman filter. Knight and Ning (2005) use the empirical characteristic function and the method of generalized moments. Strickland et al. (2003) use Bayesian estimation with a Markov chain Monte Carlo algorithm. For the model with the leverage term, Feng et al. (2004) use the Monte Carlo ML method of Durbin and Koopman (2004).

The ACD and SCD models reviewed above share the property that the dynamics of higher moments of the duration process are governed by the dynamics of the conditional mean. Ghysels et al. (2004) argue that this feature is restrictive and introduce a nonlinear two factor model that disentangles the movements of the mean and of the variance of durations. Since the second factor is responsible for the variance heterogeneity, the model is named the stochastic volatility duration (SVD) model. The departure point for this model is a standard static duration model in which the durations are independently and exponentially distributed with a gamma heterogeneity, i.e.

$$x_i = \frac{U_i}{aV_i} = \frac{H(1, F_{1i})}{aH(b, F_{2i})}, \quad (30)$$

where U_i and V_i are two independent variables with distributions $\text{gamma}(1,1)$ (i.e. exponential) and $\text{gamma}(b, b)$, respectively. The last ratio in (30) uses two independent Gaussian factors F_{1i} and F_{2i} , and $H(b, F) = G(b, \varphi(F))$, where $G(b, \cdot)$ is the quantile function of the $\text{gamma}(b, b)$ distribution and $\varphi(\cdot)$ the cdf of the standard normal. Ghysels et al. (2004) extend this model to a dynamic setup through a VAR model for the two underlying Gaussian factors. Estimation is relatively difficult and requires simulation methods.

3.4 Applications

ACD models can be used to estimate and predict the intra-day volatility of returns from the intensity of price durations. As shown by Engle and Russell (1998), a price intensity is closely linked to the instantaneous price change volatility. The latter is given by

$$\tilde{\sigma}^2(t|\mathcal{F}_t) := \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathbb{E} \left[\left(\frac{p(t+\Delta) - p(t)}{p(t)} \right)^2 \middle| \mathcal{F}_t \right], \quad (31)$$

where $p(t)$ denotes the price (or midquote) at t . By denoting the counting process associated with the event times of cumulated absolute price changes of size dp by $N^{dp}(t)$, we can formulate (31) in terms of the intensity function of the process of dp -price changes. Then, the dp -price change instantaneous volatility can be computed as

$$\begin{aligned} \tilde{\sigma}_{(dp)}^2(t|\mathcal{F}_t) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \Pr [|p(t+\Delta) - p(t)| \geq dp | \mathcal{F}_t] \cdot \left[\frac{dp}{p(t)} \right]^2 \\ &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \Pr [(N^{dp}(t+\Delta) - N^{dp}(t)) > 0 | \mathcal{F}_t] \cdot \left[\frac{dp}{p(t)} \right]^2 \\ &:= \lambda^{dp}(t|\mathcal{F}_t) \cdot \left[\frac{dp}{p(t)} \right]^2, \end{aligned} \quad (32)$$

where $\lambda^{dp}(t; \mathcal{F}_t)$ denotes the corresponding dp -price change intensity. Hence, using (14), given an estimate of the parameters of an ACD model, one can estimate or predict the instantaneous volatility of the price process $p(t)$ at any time point. Giot (2005) compares these estimates with usual GARCH based estimates obtained by interpolating the prices on a grid of regularly spaced time points. He finds that GARCH based predictions are better measures of risk than ACD based ones in a Value-at-Risk (VaR) evaluation study.

ACD and related models have been typically used to test implications of asymmetric information models of price formation. For example, the model of Easley and O'Hara (1992) implies that the number of transactions influences the price process through information based clustering of transactions. Then, including lags as well as expectations of the trading intensity as explanatory variables in a model for the price process allows to test such theoretical predictions. For a variety of different applications in market microstructure research, see Engle and Russell (1998), Engle (2000), Bauwens and Giot (2000), Engle and Lunde (2003), and Hafner (2005) among others. Several authors have combined an ACD model with a model for the marks of a financial point process. The idea is generally to model the duration process by an ACD model, and conditionally on the durations, to model the process of marks. Bauwens and Giot (2003) model the direction of the price change between two consecutive trades by formulating a competing risks model, where the direction of

the price movement is triggered by a Bernoulli process. Then, the parameters of the ACD process depend on the direction of the previous price change, leading to an asymmetric ACD model. A related type of competing risks model is specified by Bisière and Kamionka (2000). Prigent et al. (2001) use a similar model for option pricing. Russell and Engle (2005) develop an autoregressive conditional multinomial model to simultaneously model the time between trades and the dynamic evolution of (discrete) price changes.

A related string of the literature studies the interaction between the trading intensity and the trade-to-trade return volatility. Engle (2000) augments a GARCH equation for returns per time by the impact of the inverse of the observed and expected durations (x_i and Ψ_i), and of the surprise x_i/Ψ_i . A decrease in x_i or Ψ_i has a positive impact on volatility while the surprise has the reverse impact. Dionne et al. (2005) use a related model to compute an intraday VaR. Ghysels and Jasiak (1998) and Grammig and Wellner (2002) study a GARCH process for trade-to-trade returns with time-varying parameters which are triggered by the trading intensity. Meddahi et al. (2006) derive a discrete time GARCH model for irregularly spaced data from a continuous time volatility process and compare it to the ACD-GARCH models by Engle (2000) and Ghysels and Jasiak (1998).

4 Dynamic Intensity Models

In this section, we review the most important types of dynamic intensity models which are applied to model financial point processes. The class of Hawkes models and extensions thereof are discussed in Section 4.1. In Section 4.2, we survey autoregressive conditional intensity models and stochastic conditional intensity models. Statistical inference for intensity models is illustrated in Section 4.3, whereas the most important applications in the recent literature are briefly discussed in Section 4.4.

4.1 Hawkes Processes

Hawkes processes originate from the statistical literature in seismology and are used to model the occurrence of earthquakes, see e.g. Vere-Jones (1970), Vere-Jones and Ozaki (1982), and Ogata (1988) among others. Bowsher (2002) was the first applying Hawkes models to financial point processes. As explained in Section 3.2, Hawkes processes belong to the class of self-exciting processes, where the intensity is driven by a weighted function of the time distance to previous points of the process. A general class of univariate Hawkes processes is given by

$$\lambda(t) = \varphi \left(\mu(t) + \sum_{t_i < t} w(t - t_i) \right), \quad (33)$$

where φ denotes a possibly nonlinear function, $\mu(t)$ is a deterministic function of time, and $w(s)$ denotes a weight function. If $\varphi : \mathbb{R} \rightarrow \mathbb{R}^+$, we obtain the class

of nonlinear Hawkes processes considered by Brémaud and Massoulié (1996). In this case, $\mu(t)$ and $w(t)$ can take negative values since the transformation $\varphi(\cdot)$ preserves the non-negativity of the process. Such a specification is useful whenever the intensity may be negatively affected by the process history or covariates. For instance, in the context of financial duration processes, $\mu(t)$ can be parameterized as a function of covariates. Stability conditions for nonlinear Hawkes processes are derived by Brémaud and Massoulié (1996). For the special case where φ is a linear function, we obtain the class of linear Hawkes processes originally considered by Hawkes (1971). They are analytically and computationally more tractable than their nonlinear counterparts, however, they require $\mu(t) > 0$ and $w(t) > 0$ in order to ensure non-negativity.

As pointed out by Hawkes and Oakes (1974), linear self-exciting processes can be viewed as clusters of Poisson processes. Then, each event is one of two types: an immigrant process or an offspring process. The immigrants follow a Poisson process and define the centers of so-called Poisson clusters. If we condition on the arrival time, say t_i , of an immigrant, then independently of the previous history, t_i is the center of a Poisson process, $\mathcal{Y}(t_i)$, of offspring on (t_i, ∞) with intensity function $\lambda_i(t) = \lambda(t - t_i)$, where λ is a non-negative function. The process $\mathcal{Y}(t_i)$ defines the first generation offspring process with respect to t_i . Furthermore, if we condition on the process $\mathcal{Y}(t_i)$, then each of the events in $\mathcal{Y}(t_i)$, say t_j , generates a Poisson process with intensity $\lambda_j(t) = \lambda(t - t_j)$. These independent Poisson processes build the second generation of offspring with respect to t_i . Similarly, further generations arise. The set of all offspring points arising from one immigrant are called a Poisson cluster. Exploiting the branching and conditional independence structure of a (linear) Hawkes process, Møller and Rasmussen (2004) develop a simulation algorithm as an alternative to the Shedler-Lewis thinning algorithm or the modified thinning algorithm by Ogata (1981) (see e.g. Daley and Vere-Jones (2003)). The immigrants and offsprings can be referred to as "main shocks" and "after shocks" respectively. This admits an interesting interpretation which is useful not only in seismology but also in high-frequency finance. Bowsher (2002), Hautsch (2004) and Large (2005) illustrate that Hawkes processes capture the dynamics in financial point processes remarkably well. This indicates that the cluster structure implied by the self-exciting nature of Hawkes processes seem to be a reasonable description of the timing structure of events on financial markets.

The most common parameterization of $w(t)$ has been suggested by Hawkes (1971) and is given by

$$w(t) = \sum_{j=1}^P \alpha_j e^{-\beta_j t}, \quad (34)$$

where $\alpha_j \geq 0$, $\beta_j > 0$ for $j = 1, \dots, P$ are model parameters, and P denotes the order of the process and is selected exogenously (or by means of information criteria). The parameters α_j are scale parameters, whereas β_j drive the

strength of the time decay. For $P > 1$, the intensity is driven by the superposition of differently parameterized exponentially decaying weighted sums of the backward times to all previous points. In order to ensure identification we impose the constraint $\beta_1 > \dots > \beta_P$. It can be shown that the stationarity of the process requires $0 < \int_0^\infty w(s)ds < 1$, which is ensured only for $\sum_{j=1}^P \alpha_j/\beta_j < 1$, see Hawkes (1971).

While (34) implies an exponential decay, the alternative parameterization

$$w(t) = \frac{H}{(t + \kappa)^p}, \quad (35)$$

with parameters H , κ , and $p > 1$ allows for a hyperbolic decay. Such weight functions are typically applied in seismology (see e.g. Vere-Jones and Ozaki (1982) and Ogata (1988)) and allow to capture long range dependence. Since financial duration processes also tend to reveal long memory behavior (see Jasiak (1998)), specification (35) might be an interesting specification in financial applications.

Multivariate Hawkes models are obtained by a generalization of (33). Then, $\lambda(t)$ is given by the $(K \times 1)$ -vector $\lambda(t) = (\lambda^1(t), \dots, \lambda^K(t))'$ with

$$\lambda^k(t) = \varphi \left(\mu^k(t) + \sum_{r=1}^K \sum_{t_i^r < t} w_r^k(t - t_i^r) \right), \quad (36)$$

where $w_r^k(s)$ is a k -type weight function of the backward time to all r -type events. Using an exponential decay function, Hawkes (1971) suggests to parameterize $w_r^k(s)$ as

$$w_r^k(t) = \sum_{j=1}^P \alpha_{r,j}^k e^{-\beta_{r,j}^k t}, \quad (37)$$

where $\alpha_{r,j}^k \geq 0$ and $\beta_{r,1}^k > \dots > \beta_{r,P}^k > 0$ drive the influence of the time distance to past r -type events on the k -type intensity. Thus, in the multivariate case, $\lambda^k(t)$ depends not only on the distance to all k -type points, but also on the distance to all other points of the pooled process. Hawkes (1971) provides a set of linear parameter restrictions ensuring the stationarity of the process.

Bowsher (2002) proposes a generalization of the Hawkes model which allows to model point processes which are interrupted by time periods where no activity takes place. In high-frequency financial time series these effects occur because of trading breaks due to trading halts, nights, weekends or holidays. In order to account for such effects, Bowsher proposes to remove all non-activity periods and to concatenate consecutive activity periods by a spill-over function. Then, the univariate so-called generalized Hawkes model is given by

$$\lambda(t) = \mu(t) + \sum_{j=1}^P \tilde{\lambda}_j(t), \quad (38)$$

where $\tilde{\lambda}_j(0) = 0$ and

$$\tilde{\lambda}_j(t) = \pi_j \tilde{\lambda}_j(\tau_{d-1}) e^{-\rho_j(t-\tau_{d-1})} + \sum_{\tau_{d-1} < t_i \leq t} \alpha_j e^{-\beta_j(t-t_i)} \quad (39)$$

for $\tau_{d-1} < t \leq \tau_d$, where $\alpha_j \geq 0$, $\beta_1 > \dots > \beta_P > 0$, $\pi_j \geq 0$, $\rho_j > 0$ are model parameters and τ_d with $d = 0, 1, 2, \dots$ represents the end of the d -th trading period in the concatenated process. The first term in (39) captures the intensity spillover between the individual trading periods driven by the parameters π_j and ρ_j . Similarly, a multivariate version of the generalized Hawkes process is given by $\lambda(t) = (\lambda^1(t), \dots, \lambda^K(t))'$, with

$$\begin{aligned} \lambda^k(t) &= \mu^k(t) + \sum_{r=1}^K \sum_{j=1}^P \tilde{\lambda}_{r,j}^k(t), \quad (40) \\ \tilde{\lambda}_{r,j}^k(t) &= \pi_{r,j}^k \tilde{\lambda}_{r,j}^k(\tau_{d-1}) e^{-\rho_{r,j}^k(t-\tau_{d-1})} + \sum_{\tau_{d-1} < t_i \leq t} \alpha_{r,j}^k e^{-\beta_{r,j}^k(t-t_i)} \quad (41) \end{aligned}$$

for $k = 1, \dots, K$ with $\alpha_{r,j}^k \geq 0$, $\beta_{r,1}^k > \dots > \beta_{r,P}^k > 0$, $\pi_{r,j}^k \geq 0$ and $\rho_{r,j}^k > 0$. Bowsler shows that the generalized Hawkes process nests the linear Hawkes model if $\pi_{r,j}^k = 1$ and $\rho_{r,j}^k = \beta_{r,j}^k$.

4.2 Autoregressive Conditional Intensity Processes

Autoregressive conditional intensity (ACI) models constitute an alternative class of intensity models which allow to capture serial dependencies in point processes. The key idea of the ACI model introduced by Russell (1999) is to model the intensity function in terms of an autoregressive process which is updated by past realizations of the integrated intensity. Let $\lambda(t) = (\lambda^1(t), \dots, \lambda^K(t))'$. Then, Russell proposes to specify $\lambda^k(t)$ in terms of a proportional intensity structure given by

$$\lambda^k(t) = \Phi_{\tilde{N}(t)+1}^k \lambda_0^k(t) s^k(t), \quad k = 1, \dots, K, \quad (42)$$

where $\Phi_{\tilde{N}(t)+1}^k$ captures the dynamic structure, $\lambda_0^k(t)$ is a baseline intensity component capturing the (deterministic) evolution of the intensity between two consecutive points and $s^k(t)$ denotes a deterministic function of t capturing, for instance, possible seasonality effects. The function $\Phi_{\tilde{N}(t)}^k$ is indexed by the left-continuous counting function and is updated instantaneously *after* the arrival of a new point. Hence, Φ_i^k is constant for $t_{i-1} < t \leq t_i$. Then, the evolution of the intensity function between two consecutive arrival times is triggered by $\lambda_0^k(t)$ and $s^k(t)$.

In order to ensure the non-negativity of the process, the dynamic component Φ_i^k is specified in log-linear form, i.e.

$$\Phi_i^k = \exp\left(\tilde{\Phi}_i^k + z'_{i-1} \gamma^k\right), \quad (43)$$

where z_i denotes a vector of explanatory variables observed at arrival time t_i and γ^k the corresponding parameter vector. Define ε_i as a (scalar) innovation

term which is computed from the integrated intensity function associated with the most recently observed process, i.e.

$$\varepsilon_i = \sum_{k=1}^K \left(1 - \int_{t_{N^k(t_i)-1}^k}^{t_{N^k(t_i)}^k} \lambda^k(s; \mathcal{F}_s) ds \right) y_i^k, \quad (44)$$

where y_i^k defines an indicator variable that takes the value 1 if the i -th point of the pooled process is of type k . Using the random time change argument presented in Section 2.4, ε_i corresponds to a random mixture of i.i.d. centered standard exponential variates and thus is itself an i.i.d. zero mean random variable. Then, the $(K \times 1)$ vector $\tilde{\Phi}_i = (\tilde{\Phi}_i^1, \dots, \tilde{\Phi}_i^K)'$ is parameterized as

$$\tilde{\Phi}_i = \sum_{k=1}^K \left(A^k \varepsilon_{i-1} + B^k \tilde{\Phi}_{i-1} \right) y_{i-1}^k, \quad (45)$$

where $A^k = \{a_j^k\}$ denotes a $(K \times 1)$ innovation parameter vector and $B^k = \{b_{ij}^k\}$ is a $(K \times K)$ matrix of persistence parameters. Hence, the fundamental principle of the ACI model is that at each event t_i all K processes are updated by the realization of the integrated intensity with respect to the most recent process, where the impact of the innovation on the K processes can be different and also varies with the type of the most recent point. As suggested by Bowsher (2002), an alternative specification of the ACI innovation term might be $\tilde{\varepsilon}_i = 1 - \Lambda(t_{i-1}, t_i)$, where $\Lambda(t_{i-1}, t_i) := \sum_{k=1}^K \Lambda^k(t_{i-1}, t_i)$ denotes the integrated intensity of the pooled process computed between the two most recent points. Following the arguments above, $\tilde{\varepsilon}_i$ is also a zero mean i.i.d. innovation term. Because of the regime-switching nature of the persistence matrix, the derivation of stationarity conditions is difficult. However, a sufficient (but not necessary) condition is that the eigenvalues of the matrices B^k for all $k = 1, \dots, K$ lie inside the unit circle.

As suggested by Hautsch (2004), the baseline intensity function $\lambda_0^k(t)$ can be specified as the product of K different Burr hazard rates, i.e.

$$\lambda_0^k(t) = \exp(\omega^k) \prod_{r=1}^K \frac{x^r(t) p_r^s - 1}{1 + \eta_r^s x^r(t) p_r^s}, \quad (p_r^s > 0, \eta_r^s \geq 0). \quad (46)$$

According to this specification $\lambda^k(t)$ is driven not only by the k -type backward recurrence time but also by the time distance to the most recent point in all other processes $r = 1, \dots, K$ with $r \neq k$. A special case occurs when $p_r^s = 1$ and $\eta_r^s = 0, \forall r \neq s$. Then, the k -th process is affected only by its own backward recurrence time.

Finally, $s^k(t)$ is typically specified as a spline function in order to capture intraday seasonalities. A simple parameterization which is used in most studies is given by a linear spline function of the form $s^k(t) = 1 + \sum_{j=1}^S \nu_j^k (t - \tau_j)$.

$\mathbb{1}_{\{t > \tau_j\}}$, where τ_j , $j = 1 \dots, S$, denote S nodes within a trading period and ν_j the corresponding parameters. A more flexible parameterization is e.g. given by a flexible Fourier form (Gallant (1981)) as used by Andersen and Bollerslev (1998) or Gerhard and Hautsch (2002) among others.

A generalization of the ACI model has been proposed by Bauwens and Hautsch (2006). The key idea is that the multivariate intensity function $\lambda(t) = (\lambda^1(t), \dots, \lambda^K(t))'$ is driven not only by the observable history of the process but also by a common component. The latter may be considered as a way to capture the unobservable general information flow in a financial market. Such a setting turns out to be useful for the modelling of high-dimensional point processes which are driven by an unobservable common random process. By assuming the existence of a common unobservable factor $\lambda^*(t)$ following a pre-assigned structure in the spirit of a doubly stochastic Poisson process (see Section 2.3), we define the internal (unobservable) history of $\lambda^*(t)$ as \mathcal{F}_t^* . Then, we assume that $\lambda(t)$ is adapted to the filtration $\mathcal{F}_t := \sigma(\mathcal{F}_t^o \cup \mathcal{F}_t^*)$, where \mathcal{F}_t^o denotes some *observable* filtration. Then, the so-called stochastic conditional intensity (SCI) model is given by

$$\lambda^k(t) = \lambda^{o,k}(t) \left(\lambda_{\tilde{N}(t)+1}^* \right)^{\sigma_k^*}, \quad (47)$$

where $\lambda_{\tilde{N}(t)+1}^* := \lambda^*(t_{\tilde{N}(t)+1})$ denotes the common latent component which is updated at each point of the (pooled) process $\{t_i\}_{i \in \{1, \dots, n\}}$. The direction and magnitude of the process-specific impact of λ^* is driven by the parameters σ_k^* . The process-specific function $\lambda^{o,k}(t) := \lambda^{o,k}(t | \mathcal{F}_t^o)$ denotes a conditionally deterministic idiosyncratic k -type intensity component given the *observable* history, \mathcal{F}_t^o .

Bauwens and Hautsch assume that λ_i^* has left-continuous sample paths with right-hand limits and in logarithm is the zero mean AR(1) process given by

$$\ln \lambda_i^* = a^* \ln \lambda_{i-1}^* + u_i^*, \quad u_i^* \sim i.i.d. N(0, 1). \quad (48)$$

Because of the symmetry of the distribution of $\ln \lambda_i^*$, Bauwens and Hautsch impose an identification assumption which restricts the sign of one of the scaling parameters σ_k^* . The observation-driven component $\lambda^{o,k}(t)$ is specified in terms of an ACI parameterization as described above. However, in contrast to the basic ACI model, in the SCI model, the innovation term is computed based on the *observable* history of the process, i.e.

$$\varepsilon_i = \sum_{k=1}^K \left\{ -\varpi - \ln \Lambda^{o,k} \left(t_{N^k(t_i)-1}^k, t_{N^k(t_i)}^k \right) \right\} y_i^k, \quad (49)$$

where ϖ denotes Euler's constant, $\varpi = 0.5772$, and $\Lambda^{o,k} (t_{i-1}^k, t_i^k)$ is given by

$$\Lambda^{o,k} (t_{i-1}^k, t_i^k) := \sum_{j=N(t_{i-1}^k)}^{N(t_i^k)-1} \int_{t_j}^{t_{j+1}} \lambda^{o,k}(u) du$$

$$= \sum_{j=N(t_{i-1}^k)}^{N(t_i^k)-1} (\lambda_j^*)^{-\sigma_k^*} \Lambda^k(t_j, t_{j+1}) \quad (50)$$

corresponding to the sum of (piecewise) integrated k -type intensities which are observed through the duration spell and are standardized by the corresponding (scaled) realizations of the latent component. This specification ensures that ε_i can be computed exclusively based on past observables implying a distinct separation between the observation-driven and the parameter-driven components of the model. Bauwens and Hautsch (2006) analyze the probabilistic properties of the model and illustrate that the SCI model allows for a wide range of (cross-)autocorrelation structures in multivariate point processes. In an application to a multivariate process of price intensities, they find that the latent component captures a substantial part of the cross-dependences between the individual processes resulting in a quite parsimonious model. An extension of the SCI model to the case of multiple states is proposed by Koopman et al. (2005) and is applied to the modelling of credit rating transitions.

4.3 Statistical Inference

Karr (1991) shows that valid statistical inference can be performed based on the intensity function solely, see Theorem 5.2. in Karr (1991) or Bowsher (2002). Assume a K -variate point process $N(t) = \{N^k(t)\}_{k=1}^K$ on $(0, T]$ with $0 < T < \infty$, and the existence of a K -variate \mathcal{F}_t -predictable process $\lambda(t)$ that depends on the parameters θ . Then, it can be shown that a genuine log likelihood function is given by

$$\ln \mathcal{L}(\theta; \{N(t)\}_{t \in (0, T]}) = \sum_{k=1}^K \left[\int_0^T (1 - \lambda^k(s)) ds + \int_{(0, T]} \ln \lambda^k(s) dN^k(s) \right],$$

which can be alternatively computed by

$$\ln \mathcal{L}(\theta; \{N(t)\}_{t \in (0, T]}) = \sum_{i=1}^n \sum_{k=1}^K (-\Lambda^k(t_{i-1}, t_i)) + y_i^k \ln [\lambda^k(t_i)] + TK. \quad (51)$$

Note that (51) differs from the standard log likelihood function of duration models by the additive (integrating) constant TK which can be ignored for ML estimation. By applying the so-called exponential formula (Yashin and Arjas (1988)), the relation between the integrated intensity function and the conditional survivor function is given by

$$S(x_i | \mathcal{F}_{t_{i-1} + x_i}) = \exp(-\Lambda(t_{i-1}, t_i)), \quad (52)$$

which is the continuous counterpart to the well-known relation between the survivor function and the hazard rate, $S(x_i) = \exp(-\int_0^{x_i} h(u) du)$. Hence, by

ignoring the term TK , (51) corresponds to the sum of the conditional survivor function and the conditional intensity function. However, according to Yashin and Arjas (1988), the exponential formula (52) is only valid if $S(x_i|\mathcal{F}_{t_{i-1}+x_i})$ is absolutely continuous in x_i , which excludes jumps of the conditional survivor function induced by changes of the information set during a spell. Therefore, in a continuous, dynamic setting, the interpretation of $\exp(-\Lambda(t_{i-1}, t_i))$ as a survivor function should be done with caution.

The evaluation of (51) for a Hawkes model is straightforward. In the case of an exponential decay function, the resulting log likelihood function can be even computed in a recursive way (see e.g. Bowsher (2002)). An important advantage of Hawkes processes is that the individual intensities $\lambda^k(t)$ do not have parameters in common and the parameter vector can be expressed as $\theta = (\theta^1, \dots, \theta^K)$, where θ^k denotes the parameters associated with the k -type intensity component. Given that the parameters are variation free, the log likelihood function can be computed as $\ln \mathcal{L}(\theta; \{N(t)\}_{t \in (0, T]}) = \sum_{k=1}^K l^k(\theta^k)$ and can be maximized by maximizing the individual k -type components $l^k(\theta^k)$ separately. This facilitates the estimation particularly when K is large. In contrast, ACI models require to maximize the log likelihood function with respect to all the parameters jointly. This is due to the fact that the ACI innovations are based on the integrated intensities which depend on all individual parameters. The estimation of SCI models is computationally even more demanding since the latent factor has to be integrated out resulting in a n -dimensional integral. Bauwens and Hautsch (2006) suggest to evaluate the likelihood function numerically using the efficient importance sampling procedure introduced by Richard (1998). Regularity conditions for the maximum likelihood estimation of stationary simple point processes are established by Ogata (1981). For more details, see Bowsher (2002).

Diagnostics for intensity based point process models can be performed by exploiting the stochastic properties of compensators (see Bowsher (2002)) and integrated intensities given in Section 2.4. The model goodness-of-fit can be straightforwardly evaluated through the estimated integrated intensities of the K individual processes, $e_{i,1}^k := \hat{\Lambda}^k(t_{i-1}^k, t_i^k)$, the integrated intensity of the pooled process $e_{i,2} := \hat{\Lambda}(t_{i-1}, t_i) = \sum_{k=1}^K \hat{\Lambda}^k(t_{i-1}, t_i)$, or of the (non-centered) ACI residuals $e_{i,3} := \sum_{k=1}^K \left(\hat{\Lambda}^k(t_{i-1}^k, t_i^k) \right) y_i^k$. Under correct model specification, all three types of residuals must be i.i.d. standard exponential. Then, model evaluation is done by testing the dynamic and distributional properties. The dynamic properties are easily evaluated with Portmanteau statistics or tests against independence such as proposed by Brock et al. (1996). The distributional properties can be evaluated by using Engle and Russell's (1998) test against excess dispersion (see Section 3.2). Other alternatives are goodness-of-fit tests based on the probability integral transform (PIT) as employed for diagnostics of ACD models by Bauwens et al. (2004).

4.4 Applications

For financial point processes, dynamic intensity models are primarily applied in multivariate frameworks or whenever a continuous-time setting is particularly required, like, for instance, in order to allow for time-varying covariates, i.e. covariates that vary between two consecutive points of the pooled process. For sake of brevity, the case of time-varying covariates was not explicitly considered in the previous presentation. Basically, all event times associated with (discrete) changes of time-varying covariates can be treated as another point process that is not explicitly modelled. However, at each event time of the covariate process, the intensities associated with all other processes have to be updated. This requires a piecewise computation of the corresponding integrated intensities. For an extension of the ACI model to the case of time-varying covariates, see Hall and Hautsch (2006b).

One string of applications focusses on the modelling of trading intensities of different types of orders in limit order books. Hall and Hautsch (2006b) apply a bivariate ACI model to study the intensities of buy and sell transactions in the electronic limit order book market of the Australian Stock Exchange (ASX). The buy and sell intensities are specified to depend on time-varying covariates capturing the state of the market. On the basis of the buy and sell intensities, denoted by $\lambda^B(t)$ and $\lambda^S(t)$, Hall and Hautsch propose a measure of the continuous net buy pressure defined by $\Delta^B(t) := \ln \lambda^B(t) - \ln \lambda^S(t)$. Because of the log-linear structure of the ACI model, the marginal change of $\Delta^B(t)$ induced by a change of the covariates is computed as $\gamma^B - \gamma^S$, where γ^B and γ^S denote the coefficients associated with covariates affecting the buy and sell intensity, respectively (see eq. (43)). Hall and Hautsch (2006a) study the determinants of order aggressiveness and traders' order submission strategy at the ASX by applying a six-dimensional ACI model to study the arrival rates of aggressive market orders, limit orders as well as cancellations on both sides of the market. In a related paper, Large (2005) studies the resiliency of an electronic limit order book by modelling the processes of orders and cancellations on the London Stock Exchange using a ten-dimensional Hawkes process. Finally, Russell (1999) analyzes the dynamic interdependences between the supply and demand for liquidity by modelling transaction and limit order arrival times at the NYSE using a bivariate ACI model.

Another branch of the literature focusses on the modelling of the instantaneous price change volatility which is estimated on the basis of price durations, see (32) in Section 3.4. This relation is used by Bauwens and Hautsch (2006) to study the interdependence between instantaneous price change volatilities of several blue chip stocks traded at the New York Stock Exchange (NYSE) using a SCI model. In this setting, they find a strong evidence for the existence of a common latent component as a major driving force of the instantaneous volatilities on the market. In a different framework, Bowsheer (2002) analyzes the two-way interaction of trades and quote changes using a two-dimensional generalized Hawkes process.

References

- Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9:385–426.
- Andersen, T. G. and Bollerslev, T. (1998). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance*, 53:219–265.
- Bauwens, L. and Giot, P. (2000). The logarithmic ACD model: An application to the Bid/Ask quote process of two NYSE stocks. *Annales d'Economie et de Statistique*, 60:117–149.
- Bauwens, L. and Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity*. Kluwer Academic Publishers, Boston, Dordrecht, London.
- Bauwens, L. and Giot, P. (2003). Asymmetric ACD models: Introducing price information in ACD models with a two state transition model. *Empirical Economics*, 28:1.
- Bauwens, L., Giot, P., Grammig, J., and Veredas, D. (2004). A comparison of financial duration models via density forecasts. *International Journal of Forecasting*, 20:589–609.
- Bauwens, L. and Hautsch, N. (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics*, 4:450–493.
- Bauwens, L., Preminger, A., and Rombouts, J. V. (2006a). Regime switching ACD models. Forthcoming CORE Discussion Paper, Louvain-La-Neuve.
- Bauwens, L., Preminger, A., and Rombouts, J. V. (2006b). Regime switching GARCH models. CORE Discussion Paper 2006/11, Louvain-La-Neuve.
- Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: A latent factor model for the analysis of financial durations. *Journal of Econometrics*, 119:381–412.
- Bisière, C. and Kamionka, T. (2000). Timing of orders, order aggressiveness and the order book at the Paris Bourse. *Annales d'Economie et de Statistique*, 60:43–72.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.
- Bollerslev, T. and Wooldridge, J. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*, 11:143–172.
- Bowsher, C. G. (2002). Modelling security markets in continuous time: Intensity based, multivariate point process models. Economics Discussion Paper No. 2002-W22, Nuffield College, Oxford.
- Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability*, 24:1563–1588.
- Brock, W., Scheinkman, W., Scheinkman, J., and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15:197–235.
- Brown, T. C. and Nair, M. G. (1988). A simple proof of the multivariate random time change theorem for point processes. *Journal of Applied Probability*, 25:210–214.
- Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18(1):17–39.
- Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62:269.

- Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, volume 1. Springer, New York.
- De Luca, G. and Gallo, G. (2004). Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics and Econometrics*, 8 (2). Downloadable under <http://www.bepress.com/snde/vol8/iss2/art8>.
- De Luca, G. and Gallo, G. (2006). Time-varying mixing weights in mixture autoregressive conditional duration models. Manuscript, University of Florence.
- De Luca, G. and Zuccolotto, P. (2003). Finite and infinite mixtures for financial durations. *Metron*, 61:431–455.
- Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts, with applications to financial risk management. *International Economic Review*, 39:863–883.
- Dionne, G., Duchesne, P., and Pacurara, M. (2005). Intraday Value at Risk (IVaR) using tick-by-tick data with application to the Toronto Stock Exchange. Mimeo, HEC Montréal.
- Drost, F. C. and Werker, B. J. M. (2004). Semiparametric duration models. *Journal of Business and Economic Statistics*, 22:40–50.
- Dufour, A. and Engle, R. F. (2000). The ACD model: Predictability of the time between consecutive trades. Working Paper, ISMA Centre, University of Reading.
- Durbin, J. and Koopman, S. (2004). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, 84:669–684.
- Easley, D. and O'Hara, M. (1992). Time and process of security price adjustment. *The Journal of Finance*, 47:577–605.
- Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68, 1:1–22.
- Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, 17:425–446.
- Engle, R. F. and Lunde, A. (2003). Trades and quotes: A bivariate point process. *Journal of Financial Econometrics*, 11:159–188.
- Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48:1749–1778.
- Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66:1127–1162.
- Engle, R. F. and Russell, J. R. (2005). *Analysis of High Frequency Financial Data*. Handbook of Financial Econometrics. Yacine Ait-Sahalia and Lars Hansen (eds.), North-Holland.
- Feng, D., Jiang, G. J., and Song, P. X.-K. (2004). Stochastic conditional duration models with 'leverage effect' for financial transaction data. *Journal of Financial Econometrics*, 2:390–421.
- Fernandes, M. and Grammig, J. (2005). Non-parametric specification tests for conditional duration models. *Journal of Econometrics*, 127:35–68.
- Fernandes, M. and Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, 130:1–23.
- Gallant, R. A. (1981). On the bias in flexible functional forms and an essential unbiased form: The Fourier flexible form. *Journal of Econometrics*, 15:211–245.
- Gerhard, F. and Hautsch, N. (2002). Volatility estimation on the basis of price intensities. *Journal of Empirical Finance*, 9:57–89.

- Gerhard, F. and Hautsch, N. (2005). Semiparametric autoregressive proportional intensity models. Working Paper, University of Copenhagen, revised version of Working Paper W2, Nuffield College, Oxford.
- Ghysels, E., Gouriéroux, C., and Jasiak, J. (2004). Stochastic volatility duration models. *Journal of Econometrics*, 119:413–433.
- Ghysels, E. and Jasiak, J. (1998). GARCH for irregularly spaced financial data: The ACD-GARCH model. *Studies in Nonlinear Dynamics and Econometrics*, 2:133–149.
- Giesecke, K. and Tomecek, P. (2005). Dependent events and changes of time. Working Paper, Cornell University.
- Giot, P. (2005). Market risk models for intraday data. *European Journal of Finance*, 11:187–212.
- Grammig, J. and Wellner, M. (2002). Modeling the interdependence of volatility and inter-transaction duration process. *Journal of Econometrics*, 106:369–400.
- Hafner, C. M. (2005). Durations, volume and the prediction of financial returns in transaction time. *Quantitative Finance*, 5.
- Hall, A. D. and Hautsch, N. (2006a). Modelling the buy and sell intensity in a limit order book market. *Journal of Financial Markets*, forthcoming.
- Hall, A. D. and Hautsch, N. (2006b). Order aggressiveness and order book dynamics. *Empirical Economics*, 30:973–1005.
- Hamilton, J. D. and Jordà, O. (2002). A model of the federal funds rate target. *Journal of Political Economy*, 110:1135–1167.
- Han, A. and Hausman, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5:1–28.
- Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance*, 46:179–207.
- Hautsch, N. (2003). Assessing the risk of liquidity suppliers on the basis of excess demand intensities. *Journal of Financial Econometrics*, 1:189–215.
- Hautsch, N. (2004). *Modelling Irregularly Spaced Financial Data*. Springer, Berlin.
- Hautsch, N. (2006). Testing the conditional mean function of autoregressive conditional duration models. Working Paper, Department of Economics, University of Copenhagen.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90.
- Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11:493–503.
- Heinen, A. and Rengifo, E. (2003). Multivariate autoregressive modelling of time series count data using copulas. Revision of CORE Discussion Paper 2003/25, forthcoming in *Journal of Empirical Finance*.
- Henneke, J., Rachev, S., and Fabozzi, F. (2006). MCMC based estimation of Markov switching ARMA-GARCH models. Manuscript.
- Hujer, R. and Vuletic, S. (2005a). Econometric analysis of financial trade processes by discrete mixture duration models. Available at <http://ssrn.com/abstract=766664>.
- Hujer, R. and Vuletic, S. (2005b). The regime switching ACD framework: The use of the comprehensive family of distributions. Available at <http://ssrn.com/abstract=766684>.

- Hujer, R., Vuletic, S., and Kokot, S. (2002). The Markov switching ACD model. Finance and Accounting Working Paper 90, Johann Wolfgang Goethe-University, Frankfurt. Available at SSRN: <http://ssrn.com/abstract=332381>.
- Jasiak, J. (1998). Persistence in intratrade durations. *Finance*, 19:166–195.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley.
- Karr, A. F. (1991). *Point Processes and their Statistical Inference*. Dekker, New York.
- Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, 26:646–679.
- Knight, J. and Ning, C. (2005). Estimation of the stochastic conditional duration model via alternative methods—ecf and gmm. Mimeo, University of Western Ontario.
- Koopman, S. J., Lucas, A., and Monteiro, A. (2005). The multi-state latent factor intensity model for credit rating transitions. Discussion Paper TI2005-071/4, Tinbergen Institute.
- Lancaster, T. (1997). *The Econometric Analysis of Transition Data*. Cambridge University Press.
- Large, J. (2005). Measuring the resiliency of an electronic limit order book. Working Paper, All Souls College, University of Oxford.
- Liesenfeld, R., Nolte, I., and Pohlmeier, W. (2006). Modelling financial transaction price movements: a dynamic integer count model. *Empirical Economics*, 30:795–825.
- Meddahi, N., Renault, E., and Werker, B. J. (2006). Garch and irregularly spaced data. *Economics Letters*, 90:200–204.
- Meitz, M. and Saikkonen, P. (2004). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. SSE/EFI Working Paper Series in Economics and Finance No. 573, Stockholm School of Economics.
- Meitz, M. and Teräsvirta, T. (2006). Evaluating models of autoregressive conditional duration. *Journal of Business & Economic Statistics*, 24:104–124.
- Meyer, P. A. (1971). Démonstration simplifiée d’un théorème Knight. In *Lecture Notes in Mathematics*, volume 191, pages 191–195. Springer.
- Møller, J. and Rasmussen, J. (2004). Perfect simulation of Hawkes processes. Working Paper, Aalborg University.
- Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Journal of Econometrics*, 43:227–251.
- Ogata, Y. (1981). On Lewis’ simulation method for point processes. *IEEE Transactions of 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.
- Prigent, J., Renault, O., and Scaillet, O. (2001). *An Autoregressive Conditional Binomial Option Pricing Model*. Selected Papers from the First World Congress of the Bachelier Finance Society. Geman and Madan and Pliska and Vorst (eds), Springer Verlag, Heidelberg.
- Quoreshi, A. S. (2006). Long memory, count data, time series modelling for financial application. Umea Economic Studies 673, Department of Economics, Umea University.

- Richard, J.-F. (1998). Efficient high-dimensional Monte Carlo importance sampling. Working Paper, University of Pittsburgh.
- Russell, J. R. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data. Working Paper, University of Chicago.
- Russell, J. R. and Engle, R. F. (2005). A discrete-state continuous-time model of financial transactions prices and times: The autoregressive conditional multinomial autoregressive conditional duration model. *Journal of Business and Economic Statistics*, 23:166–180.
- Rydberg, T. H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, 1:2–25.
- Strickland, C. M., Forbes, C. S., and Martin, G. M. (2003). Bayesian analysis of the stochastic conditional duration model. Monash Econometrics and Business Statistics Working Paper 14/03, Monash University.
- Vere-Jones, D. (1970). Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society, Series B*, 32:1–62.
- Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical inference applied to earthquake data. *Annals of the Institute of Statistical Mathematics*, 34:189–207.
- Veredas, D., Rodriguez-Poo, J., and Espasa, A. (2002). On the (intradaily) seasonality, dynamics and durations zero of a financial point process. CORE DP 2002/23, Louvain-La-Neuve.
- Yashin, A. and Arjas, E. (1988). A note on random intensities and conditional survival functions. *Journal of Applied Probability*, 25:630–635.
- Zhang, M. Y., Russell, J., and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104:179–207.