UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 09 – 79
Statistics and Econometrics Series 24
December 2009

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 6249849

THE ECONOMETRICS OF RANDOMLY SPACED FINANCIAL DATA: A SURVEY

André A. Monteiro[1]

Abstract

This paper provides an introduction to the problem of modeling randomly spaced longitudinal data. Although Point Process theory was developed mostly in the sixties and early seventies, only in the nineties did this field of Probability theory attract the attention of researchers working in Financial Econometrics. The large increase, observed since, in the number of different classes of Econometric models for dealing with financial duration data, has been mostly due to the increased availability of both trade-by-trade data from equity markets and daily default and rating migration data from credit markets. This paper provides an overview of the main Econometric models available in the literature for dealing with what is sometimes called *tick data*. Additionally, a synthesis of the basic theory underlying these models is also presented. Finally, a new theorem dealing with the identifiability of latent intensity factors from point process data, jointly with a heuristic proof, is introduced.

Keywords: Tick Data; Financial Duration Models; Point Processes; Migration Models
JEL classification: C22, C32, C34, C41, G10

[1]Universidad Carlos III de Madrid, Departament of Statistics, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: andreantonio.monteiro@uc3m.es.

The Econometrics of Randomly Spaced Financial Data: A Survey

André A.V.P.B. Monteiro*
Department of Statistics, Carlos III University of Madrid

This version: November 2009

Abstract

This paper provides an introduction to the problem of modeling randomly spaced longitudinal data. Although Point Process theory was developed mostly in the sixties and early seventies, only in the nineties did this field of Probability theory attract the attention of researchers working in Financial Econometrics. The large increase, observed since, in the number of different classes of Econometric models for dealing with financial duration data has been mostly due to the increased availability of both trade-by-trade data from equity markets and daily default and rating migration data from credit markets. This paper provides an overview of the main Econometric models available in the literature for dealing with what is sometimes called *tick data*. Additionally, a synthesis of the basic theory underlying these models is also presented. Finally, a new theorem dealing with the identifiability of latent intensity factors from point process data, jointly with a heuristic proof, is introduced.

Keywords: Tick Data, Financial Duration Models, Point Processes, Migration Models
JEL classification codes: C22, C32, C34, C41, G10

* Contact details: Department of Statistics, Universidad Carlos III de Madrid, C/ Madrid 126, Getafe 28903, Spain. Tel. +34916249314 Fax. +34916249849 email: andreantonio.monteiro@uc3m.es. An early draft of this survey paper was included in my PhD Dissertation at the Free University Amsterdam.

1 Introduction

The increasing availability of (ultra-)high-frequency data, arising from financial markets, led in recent years to a dramatic growth in the literature dealing with the Econometric tools needed for handling it efficiently. The traditional methods based on fixed-length intervals of time are simply not adequate for dealing with this type of data, as they require the aggregation of the observations to the level of the corresponding time-grid. In general, the aggregation of randomly spaced event-data into one or more time-series of counts, or of some general discrete-response variable, leads to several known problems. First, if the length of the chosen observational unit of time is 'too short' then there will be many intervals where there are no observations at all. This introduces, artificially, a very specific type of heteroscedasticity in the resulting time-series. On the other hand, if one chooses a 'large' time-unit then the underlying micro structure features of the data are lost, which can be a serious problem for some analyses. Third, in many problems the duration between two well-defined successive events, or the variability in the frequency of those events, is the subject of the study, and here, clearly no aggregation should be taken. Although, for practical measuring purposes, there is always some very small basic time-unit. In the case of tick data this can be as small as 1/100th of a second.

Finally, the aggregation of point process data over some fixed-length interval of time does not allow the researcher to account for changes on time-varying covariates of interest taking place during the duration of that interval.

The alternative to fixed-interval techniques is to model the Data Generating Process (DGP) behind the successive events being recorded as a Random Point Process, eventually a multivariate, marked or generalized, non-stationary one.

The Mathematical Theory of Point Processes is a topic finding its origins in the work of the 19^{th} century French mathematician S. D. Poisson dealing with the analysis of Life tables (i.e. statistics from the distribution of the duration of human life) and mortality rates. The modern treatment of the subject, however, goes back to the works of Cox and Wold in the early fifties. During the sixties several central limit theorems concerning the superposition of Random Point Processes were proved. These theorems basically showed, for the first time, that the Poisson Point Process (one of the simpler Random Point Process models) plays a role in Point Processes theory similar to that of the Gaussian distribution in the Theory of Distributions.

Recently, Point Process theory has drawn a lot of attention from applied researchers in Financial Econometrics. The seminal work of Engle and Russell (1998), introducing the Autoregressive Conditional Duration (ACD) model, spurred an entire new stream of literature focusing on

econometric specifications for modeling Point Processes evolving with after-effects.[1] Most of these econometric specifications were developed for the analysis of financial tick data. A related, but distinct literature on *migration models* for credit risk appeared during the same period (see for example Jarrow et al., 1997, and more recently Gagliardini and Gourieroux, 2005). This paper intends to provide an unified perspective on both these streams of literature.

The focus of most of the models in this survey is placed on (ultra-)high-frequency data sets. However, there is nothing in the structure of these models that prevents their use for the analysis of lower-frequency randomly spaced data. In fact, the Point Processes approach is not exclusively motivated by high-frequency (or high-intensity[2]) data. A very prolific stream of literature on Duration Analysis (i.e. the study of the distribution of the inter-arrival times - the amount of time elapsed between two successive event-points of a point process) stemming from the path braking work of Cox (1972), in particular from his so-called Proportional Hazards model, has been widely applied in the fields of Medical Science, Biostatistics, Labor Economics and Sociology. Here, typically, the type of durations under study are better expressed in weeks, months or even years (for example the duration of the unemployment and employment spells for workers).

The different econometric point process models available in the literature can be classified according to, at least, two different criteria. Perhaps due to the influential work of Engle and Russell (1998), many econometric point process models directly specify the *forward occurrence density*[3] conditional on a particular filtration. Examples include the already mentioned ACD model (and all derived models), but also the Stochastic Conditional Duration (SCD) model of Bauwens and Veredas (2004). A more flexible alternative consists in specifying directly the intensity process associated with the point process. This approach, which was in fact the central paradigm in Duration Analysis since, at least, the already-mentioned work of Cox (1972), was recovered by Russell (1999) for his Autoregressive Conditional Intensity (ACI) model. Modeling directly the intensity of the PP is also the approach taken by Bauwens and Hautsch (2006a) for the Stochastic Conditional Intensity (SCI) model, and by Bowsher (2007) for his generalized Hawkes processes.

Econometric point process models can, however, also be classified according to a general criterion used extensively across Statistics. Models like the above-mentioned ACD, ACI and generalized Hawkes processes are good examples of *observation driven* models. This class of models assumes that, conditional on a particular observable filtration (encompassing the internal one), the distribution of the future observations is completely specified. In contrast,

[1] see definition 4.

[2] See equations (5) and (8).

[3] See definition 7.

parameter driven models explicitly acknowledge that in most real World situations, even conditioning on a rich observable filtration, the probability law governing future observations cannot be known precisely. That is, often this probability law is itself random. Parameter driven models, however, postulate that the *shape* of this probability law is known. The randomness being due to the presence of unobservable factors. These latent factors are modeled as stochastic processes inside a fully parametric setting.

Bauwens and Hautsch (2006b) provides an overview of the econometric point processes literature somewhat similar to the one contained in the current paper. There are, however, a number of distinctive features between the later and the former. In particular, in this paper the two main classes of *Transition Processes* (or generalized point processes) are also covered. These types of stochastic processes are directly relevant for the analysis of credit rating data, and are outside the scope of the survey by Bauwens and Hautsch (2006b). This paper also provides a significantly more extensive overview of the theory of point processes. This provides the reader with a deeper insight into some of the technical issues associated with each particular specification. Two new theorems focusing on random changes of the time-scale and the identification of latent intensity factors in point process data are also introduced in this paper. This paper also makes a minor contribution to the econometric literature dealing with models for (ultra-)high-frequency data by presenting an innovative and more concise vector-matrix notation for some of the models covered. This is the case for the ACI, SCI and generalized Hawkes models.[4] Finally, in this paper all point process models surveyed are explicitly classified as either observation or parameter driven. This allows readers familiar with the literature on (generalized) State Space Models to grasp more easily the technical issues connected with the estimation of the different specifications covered.

This survey paper is organized as follows. Section 2 reviews the essential concepts and results from the Theory of Point Processes and Markov and semi-Markov processes in continuous time. Sections 3 and 4 constitute the core of the paper. Here I briefly review the main Econometric models in the literature dealing with empirical Point Processes. As mentioned, the different models in the literature can be grouped in two large classes. In Section 3 the focus is placed on *observation driven* models. Parameter driven models are the focus of Section 4. Section 5 concludes.

[4] As a further minor contribution, a small technical issue connected with the specification of the SCI model is clarified. See the discussion and footnote following formula (70) in Subsection 4.2.

2 Theoretical Background

This section reviews important background concepts and results from the general theory of Point Processes (PP), Classical Duration and Event-History Analysis and from the theory of Continuous Time Markov Chains. These concepts are useful for understanding clearly the assumptions and mechanisms behind each econometric PP model. PP theory provides a structured and unified framework that accommodates many seemingly disparate models. Some authors reserve the term "duration models"only for those PP models that directly specify the forward occurrence density. According to this view, models that directly specify the (continuous-time) intensity process associated with a PP are called *intensity models*. However, this distinction is slightly artificial. Not only because the intensity process of a PP directly implies a particular forward occurrence density, but also because these so-called 'duration models' also imply a particular type of (continuous-time) intensity process - one that directly depends on the backward-recurrence time of the PP, i.e. the time since the last observed event. For this reason, I prefer to designate as "duration models"any PP model that does not explicitly assume the existence of a predetermined time-grid over which the observed number of events is counted. The later case corresponds to the so-called "count data models."In fact, bot duration and count data models are aimed at describing empirical PP. However, as mentioned in the previous section, count data models, unlike their duration counterparts, imply a loss of information due to the aggregation of the number of events over the corresponding time-grid. Therefore, in this section, the emphasis is placed more on the *duration* and *intensity* statistics rather than on the *counting* statistics associated with a PP.

2.1 Random Point Processes

In this section I provide a brief overview of some important results from the theory of real-valued Point Processes. The literature on PP theory, however, is reasonably extensive. An easily accessible and detailed treatment of the main results of the Theory of Point Processes is available in Snyder (1975), which concentrates on temporal PPs. Snyder and Miller (1991) provides a detailed treatment of more general PPs in multidimensional spaces. Karr (1991) provides a formal treatment of the subject from a Measure Theoretical perspective. A comprehensive and less demanding introduction to the subject is given in Daley and Vere-Jones (2002). Andersen et al. (1993) sets out in full mathematical detail the modern, martingale-based counting processes approach to the theory of PP.

Definition 1 ***(Random Point Process)***

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ *denote a filtered Probability Space satisfying the usual conditions (Andersen et al., 1993), a real Random Point Process*

$$\{..., T_{-1}(\omega), T_0(\omega), T_1(\omega), ..., T_n(\omega), ...\},$$

where the random variables $T_n : \Omega \to \mathbb{R}$ *satisfy* $T_{n-1}(\omega) \leq T_n(\omega)$ $\forall_{n\in\mathbb{Z},\omega\in\Omega}$*, is a function from* Ω *into the set of all nondecreasing sequences in* $\mathbb{R}$*.*

In this definition it should be noted, that for a particular $\omega \in \Omega$ the sequence $\{T_n(\omega)\}$ may actually be finite. The random event-moments T_n are in fact *stopping times* with respect to the filtration $\{\mathcal{F}_t\}$.

Definition 2 ***(Counting Process associated with a Point Process):***

Given a real Point Process $\{T_n\}_{n=-\infty}^{\infty}$ *and a subset* A *of* $\mathbb{R}$*, the associated* Counting Process $N(A)$ *is defined as the number of occurrences of the point process in the set* A*, formally:*

$$N(A) = \sum_n \mathbf{1}_A(T_n),$$

where $\mathbf{1}_A(.)$ *denotes the indicator function over the set* A*.*

Of particular relevance for modeling purposes are the special cases where $A = [t_0, t]$ and $A = [t_0, t)$ with $t_0 < t$. For a fixed $t_0 \in \mathbb{R}$, and with a slight abuse of notation, I write $N([t_0, t])$ as $N(t)$, and $N([t_0, t))$ as $\bar{N}(t)$. Letting $t \in \mathbb{R}$ vary, the stochastic process $N(t)$ has càdlàg sample paths which are piecewise constant, while those from $\bar{N}(t)$ are càglàd piecewise constant.
Consider a fixed $t_0 \in \mathbb{R}$, a random point process is said to be *non explosive* in the interval $[t_0, t]$ if $\mathrm{E}[N(t)] < \infty$. Because for every $t > t_0$ we have $\mathrm{E}[N(t) - \bar{N}(t)] \geqslant 0$, the counting process $N(t)$ is a *submartingale*, that is $\mathrm{E}[N(t)|\mathcal{F}_u] \geqslant N(u), \forall_{u,t} : t_0 \leq u < t$. The *Doob-Meyer decomposition* establishes the existence of a unique càdlàg, nondecreasing, $\mathcal{F}_t-$predictable process $\Lambda(t)$, which is the *compensator* of $N(t)$. This means the process defined as

$$M(t) = N(t) - \Lambda(t),$$

is a $\mathcal{F}_t-$martingale. Therefore, the following holds

$$\mathrm{E}[N(t) - N(u)|\mathcal{F}_u] = \mathrm{E}[\Lambda(t) - \Lambda(u)|\mathcal{F}_u], \ \forall_{u,t} : t_0 \leq u < t. \tag{1}$$

If $\Lambda(t)$ is absolutely continuous, then there is an $\mathcal{F}_t-$predictable *intensity process* $\lambda(t)$ such that

$$\Lambda(t) = \int_{t_0}^{t} \lambda(\tau)\mathrm{d}\tau. \tag{2}$$

Definition 3 ***(Orderliness):***

A Counting Process, and the underlying Point Process, are called orderly *at time $t \geqslant t_0$ if for any given $\varepsilon > 0$ there exists a $\delta(t, \varepsilon) > 0$ such that*

$$P[N([t, t+\Delta)) > 1] \leq \varepsilon P[N([t, t+\Delta)) = 1],\ \forall_\Delta : 0 < \Delta < \delta(t, \varepsilon).$$

A Point Process is orderly in an interval $[a, b]$ *if it is orderly at every point of that interval. It is* uniformly orderly *on the interval if $\delta(t, \varepsilon) = \delta(\varepsilon)$.*

Intuitively this property means that, for an orderly point process, the probability of observing more than one point in a given time-interval can be made an arbitrarily small fraction of the probability of observing one single point, provided the interval is small enough (in fact if we could rule out the possibility of $\mathrm{P}[N([t, t+\Delta)) = 1] = 0, \forall_{\Delta>0}$ and $t \geqslant t_0$ then this definition would be equivalent to simply say that $\lim_{\Delta \downarrow 0} \frac{\mathrm{P}[N([t,t+\Delta))>1]}{\mathrm{P}[N([t,t+\Delta))=1]} = 0$ and uniform orderliness means that this probability ratio converges uniformly to zero).

Definition 4 ***(Evolution without after-effects):***

A point process observed over an interval $[t_0, \infty)$ is said to evolve without after-effects if for any $t > t_0$ the realization of points over the interval $[t, \infty)$ does not depend in any way on the realization over the interval $[t_0, t)$.

In particular, this implies that the (integer) random variables $(N(u) - N(u'))$, $\forall_{u>u'}$, and $N(\tau) - N(\tau')$, $\forall_{\tau>\tau'>u}$ are independent. That is, a point process evolving without after-effects has *independent increments*.

Definition 5 ***(Poisson Processes):***

Let $\{N(t); t \geqslant t_0\}$ be the counting process associated to a point process defined on the interval $[t_0, \infty)$, this point process is called a Poisson Point Process (and $\{N(t); t \geqslant t_0\}$ a Poisson counting process) if the following conditions hold:

1. *P[$N(t_0) = 0$]=1;*

2. *for $t_0 \leqslant u \leqslant t$, the increment $N((u,t]) = N(t) - N(u)$ is Poisson distributed with parameter $\Lambda(t) - \Lambda(u)$, i.e.*

$$P[N((u,t]) = n] = \frac{(\Lambda(t) - \Lambda(u))^n \exp\left[-\left(\Lambda(t) - \Lambda(u)\right)\right]}{n!}. \tag{3}$$

   Where $\Lambda : [t_0, \infty) \to \mathbb{R}_0^+$ is an arbitrary, non-decreasing deterministic function satisfying $\Lambda(t_0) = 0$.

3. *$\{N(t); t \geqslant t_0\}$ has independent increments. That is, the number of points in non-overlapping intervals are independent random variables.*

Recalling that the mean of a Poisson distribution (and actually all its cumulants) equals the single parameter of this distribution, we can see that the general property (1) implies that the compensator of the Poisson counting process coincides with its parameter function (defined in point 2 above, therefore justifying the use of identical notation).

The Poisson Process is without doubt the single most important Random Point Process model. It can be viewed as the natural benchmark model when analyzing a particular empirical point process. Most of the point processes covered in this survey can be thought of as resulting from relaxing some of the qualitative assumptions behind the Poisson Process (summarized in theorems 1 and 2 presented further ahead in this section).

The properties of the parameter function directly affect the behavior of the Poisson point process, in particular,

- If $\Lambda(t) - \Lambda(u)$ is finite, then points do not occur *with certainty* (i.e. there is never probability one of observing at least one point) over the interval $(u,t]$ and there is also zero probability of observing an infinite number of points in that interval. Additionally, from the well-known fact that for a Poisson distribution the mean and variance coincide, we see that

$$\mathrm{E}\left[N((u,t])\right] = \mathrm{V}\left[N((u,t])\right] = \Lambda(t) - \Lambda(u), \tag{4}$$

- The points d_i of discontinuity of $\Lambda(t)$ correspond to *singular time-points* of the point process. That is, at these points there is a non-zero probability mass (equal to $1 - \exp\left[-\left(\Lambda(d_i^+) - \Lambda(d_i^-)\right)\right]$) of observing at least one event-point. More accurately still, at these predetermined time-points the number of events follows a Poisson distribution with parameter equal to the size of the discontinuity jump in $\Lambda(t)$ (i.e. $\Lambda(t_i^+) - \Lambda(t_i^-)$),

- If $\Lambda(t)$ is continuous for all $t \geqslant t_0$ then event-points do not occur at predetermined times and $\lim_{\Delta \downarrow 0} \mathrm{P}\left[N([t, t+\Delta)) = 0\right] = 1$.

Definition 6 ***(Intensity function of a Poisson process):***

The intensity function *of a Poisson process is defined at the points of differentiability of the parameter function* $\Lambda(t)$ *as its first derivative,*

$$\lambda(t) = \lim_{\Delta \to 0} \frac{\Lambda(t+\Delta) - \Lambda(t)}{\Delta}. \tag{5}$$

Because, when Λ is differentiable at t, the expected number of points in the interval $[t, t+\Delta)$ equals $\Lambda(t+\Delta) - \Lambda(t)$ this intensity function can be interpreted as the *instantaneous mean rate* at which event-points occur.
If $\Lambda(t)$ is absolutely continuous then it can be expressed as

$$\Lambda(t) = \int_{t_0}^{t} \lambda(\tau) \mathrm{d}\tau. \tag{6}$$

So far I have mostly considered the *counting statistics* associated with the Poisson Point Process.

That is, the statistics associated with the distribution of the number of event-points over arbitrary intervals of time.
Another important aspect of a Point Process are the so-called *time statistics*. These include both the statistics associated with the distributions of the sequences of inter-arrival times and with the *degree of clustering* of event-points over time.
Two closely related time-sequences can be distinguished,

- The sequence of *occurrence times* $\{T_1, T_2, ..., T_{N_t}\}$ where the event-points were recorded over the interval $[t_0, t]$,
- The sequence of *inter-arrival times* (or *durations*) $\{\tau_1, \tau_2, ..., \tau_{N_t}\}$ defined as $\tau_i = T_i - T_{i-1}$ implying $T_i = \sum_{j=0}^{i} \tau_j$, with the convention $\tau_0 = t_0$.

Clearly, both these sequences yield the same amount of information about the history of the point process over the interval $[t_0, t]$.

Definition 7 ***(Forward-Occurrence density)***

The conditional density of the next inter-arrival time (eventually the first) given the past sequence of occurrence times (eventually an empty sequence) is called the forward-occurrence density.

For the case of a Poisson Process this density is given by

$$\begin{aligned} f_{\tau_n|T_{n-1},...,T_1}(\tau \mid t_{n-1},...,t_1) &= \lambda(t_{n-1}+\tau)\exp\left[-\left(\Lambda(t_{n-1}+\tau)-\Lambda(t_{n-1})\right)\right] \\ &= \lambda(t_{n-1}+\tau)\exp\left[-\int_{t_{n-1}}^{t_{n-1}+\tau}\lambda(u)\mathrm{d}u\right] \end{aligned}$$

the last expression being valid if $\Lambda(t)$ is absolutely continuous. A special case arises if the intensity function $\lambda(t)$ is constant (this case is termed a *homogeneous Poisson point process*),

$$f_{\tau_n|T_{n-1},...,T_1}(\tau \mid t_{n-1},...,t_1) = \lambda\exp(-\lambda t),$$

so, in this special case, the forward-occurrence density is a Exponential density with mean $1/\lambda$. In what follows, two theorems giving qualitative conditions for a general point process to be a

Poisson point process are presented. The first theorem (following Khinchin, 1956) has mainly a theoretical interest. It presents the weakest set of sufficient conditions for an arbitrary point process to be a Poisson process.[5] However, these conditions are not easily relaxed for obtaining more general classes of point processes. The second theorem, on the contrary, has more an operational interest. Relaxing some of the conditions in this second theorem leads to more general classes of point processes that actually contain most of the specific processes covered in this paper.

Theorem 1 *A counting process $\{N_t; t \geqslant t_0\}$ associated to a point process is a Poisson counting process if the following qualitative conditions are met,*

1. the point process is uniformly orderly on $[t_0, t)$, for all $t > t_0$,
2. the point process evolves without after-effects,
3. points do not occur at predetermined times,
4. there is no finite sub-interval of $[t_0, t)$ where points occur with certainty,
5. $\mathrm{P}[N(t_0) = 0] = 1$.

The second set of sufficient conditions for a point process to be a Poisson process (assembled under theorem 2), unlike the one in theorem 1, does not imply the most general Poisson process

[5] These conditions are sufficient to obtain the most general (non-singular) Poisson counting process possible, one having a continuous nonnegative and non-decreasing parameter function $\Lambda(t)$, termed an inhomogeneous Poisson process

possible. However, modifying slightly these conditions, leads to a particularly important class of (non-Poisson) point processes, the so-called *Self-Exciting point processes*. We first need the concept of *conditional orderliness*, which is a stronger version of the orderliness notion already seen.

Definition 8 ***(Conditional Orderliness):***

A Counting Process, and the underlying Point Process, are Conditionally Orderly *at time* $t \geqslant t_0$ *if for any given* $\varepsilon > 0$ *there exists a* $\delta(t, \varepsilon)$ *such that*

$$P[N([t, t+\Delta)) > 1 | \mathcal{F}_{t^-}] \leqslant \varepsilon P[N([t, t+\Delta)) = 1 \mid \mathcal{F}_{t^-}], \ \forall_\Delta : 0 < \Delta < \delta(t, \varepsilon). \tag{7}$$

The process is called conditionally orderly if it is conditionally orderly for all $t \geqslant t_0$. A conditionally orderly point process is also (unconditionally) orderly but the converse is not necessarily true.

In intuitive terms this means that the orderliness of the process stays unaffected by any possible event that may occur in its past history.

A second theorem stating an alternative set of sufficient conditions for a point process to be a Poisson process is now presented (For a proof see Snyder, 1975).

Theorem 2 *A counting process* $\{N(t); t \geqslant t_0\}$*, associated with a given point process, is a Poisson counting process if it satisfies the following conditions,*

1. the point process is conditionally orderly,
2. for all $t \geqslant t_0$ the limit

$$\lambda(t) = \lim_{\Delta \downarrow 0} \frac{\mathrm{P}\left[N([t, t+\Delta)) = 1 \mid \mathcal{F}_{t^-}\right]}{\Delta}, \tag{8}$$

   exists and is a finite integrable function that depends *only* on t. Therefore, we can define $\Lambda(t) = \int_{t_0}^{t} \lambda(\tau) \mathrm{d}\tau$, which is also a finite function $\forall_{t > t_0}$,
3. $\mathrm{P}[N(t_0) = 0] = 1$,

Note that if we set $n = N(t)$ and define $T = \sum_{i=0}^{n+1} \tau_i$, where $\tau_1, \tau_2, \ldots, \tau_{n+1}$ denote the first $(n+1)$ inter-arrival times, then under condition (1) of Theorem 2, we can also interpret $\lambda(t)$ defined in (8) as

$$\lambda(t) = \lim_{\Delta \downarrow 0} \frac{\mathrm{P}\left[t \leqslant T < t + \Delta \mid T \geqslant t\right]}{\Delta}. \tag{9}$$

For any conditionally orderly point process (not necessarily a Poisson PP), equation (8) defines

the *conditional intensity process* of the PP.

In almost all applications, it is necessary to conduct estimation and inference over observed Poisson processes. The critical quantity to be computed for this purpose is the Likelihood functional at the observed realization of the Poisson PP. This one is basically the joint probability density of the observed sample path of the PP, considered as a functional of the intensity function.

2.1.1 The Likelihood of a finite sample path of a Poisson Process

Consider a realization of a Poisson counting process $N(t)$ over a given finite interval: $\{N(u); t_0 \leqslant u \leqslant t\}$ the Likelihood functional for a given intensity function λ conditional on this sample path is given by,

$$L\left(\lambda \mid \{N(u); t_0 \leqslant u \leqslant t\}\right) = \exp\left[\int_{t_0}^{t} \ln \lambda(\tau) \mathrm{d}N(\tau) - \int_{t_0}^{t} \lambda(\tau) \mathrm{d}\tau\right]. \tag{10}$$

The full set of Likelihood-based inference and testing procedures are readily available for Poisson PP.

The reason why the Poisson point process is such an important model for empirical point processes is threefold. First, as mentioned, the Poisson point process provides a baseline model whose generalizations lead to the two most important classes of point process models. These are the, already mentioned, self-exciting point processes and the class of *Doubly Stochastic Poisson processes*. Second, many empirical point processes are, in fact, adequately described by a Poisson process. This is in part due to the fact that the superposition of many independent point processes leads to a Poisson PP (see Snyder, 1975, for a treatment of some of these Poisson 'central limit' theorems). Third, a particularly important result, the *Random Time Change Theorem* (Meyer, 1971, Brown and Nair, 1988) establishes that any multivariate point (counting) process whose corresponding (multivariate) compensator is absolutely continuous and unbounded can be mapped into a set of independent homogeneous Poisson point processes each with unit intensity.

Theorem 3 ***(Random Time Change)*** *Let $(N_1(t), \ldots, N_S(t))$ be a multivariate counting process associated with S given point processes observed over $[t_0, \infty]$, with continuous, unbounded vector-compensator process $(\Lambda_1(t), \ldots, \Lambda_S(t))$. Let the corresponding S sequences of occurrence times be denoted as $\{T_n^s\}_{n=1}^{\infty}$, with $s = 1, \ldots, S$. Then the individual counting processes $\tilde{N}_s(u) = N(\Lambda_s^{-1}(u))$, whose occurrence times are given by $\{\Lambda_s(T_n^s)\}_{n=1}^{\infty}$, make up a set of independent Poisson point processes with unit intensity.*

Proof. See Brown and Nair (1988). ■

As it turns out, see Aalen and Hoem (1978) and also Andersen et al. (1993, Subsection II.5.2.2.), the type of time transformation $t \mapsto u$ employed in the Random Time Change Theorem, that is $u = \Lambda_s(t)$, can be extended to any *non-decreasing*, adapted, continuous (and therefore predictable) process $\Phi(t)$. In what follows I present one extension of the univariate Random Time Change Theorem for the class of counting processes having a multiplicative intensity process.

Theorem 4 ***(Multiplicative Intensity Univariate Random Time Change).*** *Let $N(t)$ be an $\mathcal{F}_t$-adapted counting process observed over $[t_0, \infty]$, having an absolutely continuous, unbounded, compensator process $\Lambda(t)$. Let the corresponding sequence of occurrence times be denoted as $\{T_n\}_{n=1}^{\infty}$. Assume that the corresponding intensity process can be factored as $\lambda(t) = \phi(t)\psi(t)$, where $\phi(t)$ and $\psi(t)$ are two $\mathcal{F}_t$-adapted, càglàd, non-negative processes, and define $\Phi(t) = \int_{t_0}^{t} \phi(\tau)d\tau$. Further assume that $\Phi(\infty) = \lim_{t\to\infty} \Phi(t) = \infty$.*
Then, the point process $\tilde{N}(u) = N(\Phi^{-1}(u))$, whose occurrence times are given by $\{\Phi(T_n)\}_{n=1}^{\infty}$, has the $\tilde{\mathcal{F}}_u$-predictable intensity process $\tilde{\lambda}(u) = \psi(\Phi^{-1}(u)) = \psi(t)$, where $\tilde{\mathcal{F}}_u = \mathcal{F}_{\Phi^{-1}(u)}$.

Proof. This result is in fact a corollary of Theorem 3.2 in Aalen and Hoem (1978). To see this simply set $k = 1$ in that theorem and note that $\Phi(t_0) = 0$, $\Phi(\infty) = \infty$ and the left-derivative of $\Phi(t)$, that is $\phi(t)$, is left-continuous. Therefore, the random time transformation $u = \Phi(t)$ is a *regular time change* for $N(t)$, as clearly $\phi(t) = 0$ implies $\lambda(t) = 0$. ■

Definition 9 ***(Marked Point Processes):***

A Marked Point Process is a point process that has an auxiliary random variable (called the mark) associated with every event-point. This random variable can take values on some mark set $\mathbf{M}$. *That is, we can think of a marked point process as a pair (T_n, M_n) where $\{T_n\}_{n=-\infty}^{\infty}$ is a real point process and M_n is a random variable defined over* $\mathbf{M}$ *(either a denumerable or continuous set but independent of the index n).*

As seen, a point process must be both conditionally orderly and evolving without after-effects, in order to be a Poisson process. Certain types of point processes not satisfying the orderliness requirement can be included in the class of Marked point processes. A conditionally non-orderly point process can be modeled as a marked point process where the mark represents the number of simultaneous events recorded at T_n, that is, where the mark space is $\mathbf{M} = \mathbb{N}$.
Removing the requirement that the point process evolves without after-effects leads to one of

the most important classes of dynamic PP models. Self-exciting point processes evolve *with* after-effects. This means, in the most general setting, that, at time t, any subset of the history of the process $\{N(u); t_0 \leqslant u \leqslant t\}$ can potentially influence any subset of the corresponding future $\{N(u); u > t\}$. The dependence of the future on the past is formalized through the *conditional intensity function* of the process defined in equation (8).

Definition 10 ***(Self-Exciting Point Processes):***

The conditionally orderly counting process $N(t)$ (and the underlying point process) is termed a Self-Exciting Counting (point) process if $\lambda(t)$, as defined in equation (8)*, is not merely a function of time, but instead, a stochastic process adapted to the internal filtration generated by $N(t)$.*

Again, we are interested in performing Likelihood-based estimation and inference procedures for self-exciting PP models. The important fact here is that the Likelihood functional computed for a specific realization of a general self-exciting PP is completely similar to the Likelihood functional for a (inhomogeneous) Poisson PP.

2.1.2 The Likelihood functional for Self-Exciting Point Processes

Consider a realization of a Self-Exciting counting process $N(t)$ over a given finite interval: $\{N(u); t_0 \leqslant u \leqslant t\}$ the Likelihood functional for a given *realization* of the stochastic intensity process λ conditional on this sample path is given by,

$$L\left(\lambda(\theta) \mid \{N(u); t_0 \leqslant u \leqslant t\}\right) = \exp\left[\int_{t_0}^{t} \ln \lambda(\tau;\theta) \mathrm{d}N(\tau) - \int_{t_0}^{t} \lambda(\tau;\theta) \mathrm{d}\tau\right]. \tag{11}$$

Here it is assumed that the admissible intensity processes for a specific problem are parameterized in terms of the (eventually infinite-dimensional) vector θ.

The similarity between the Poisson and the Self-Exciting cases is a consequence of the fact that the intensity process for this last class of Point Processes is a stochastic process adapted to the internal filtration of the PP. Parameter estimation can be performed through the common Maximum Likelihood (ML) methodology. In a similar, way Likelihood-based inference procedures are readily available, for example the test of two simple hypothesis:

$$H_0 \;:\; \{\lambda(\tau) = \lambda(\tau)^{(0)} \equiv \lambda(\tau;\theta_0); t_0 \leqslant \tau \leqslant t\} \;\text{ vs.} \tag{12}$$

$$H_1 \;:\; \{\lambda(\tau) = \lambda(\tau)^{(1)} \equiv \lambda(\tau;\theta_1); t_0 \leqslant \tau \leqslant t\} \tag{13}$$

can be performed using the standard Likelihood Ratio approach.

Definition 11 ***(m-Memory Self-Exciting Point Processes)***

In the general definition of self-exciting point processes the entire history of the process may influence the conditional intensity process (that is the total number of occurrences $N(t)$ and their occurrence times $T_1, T_2, ..., T_{N(t)}$). By contrast for a m-Memory Self-Exciting point process only the last m occurrence times $T_{N(t)}, T_{N(t)-1}, ..., T_{N(t)-m+1}$ *(and, eventually, the total number of occurrences $N(t)$) influence the conditional intensity process.*

A particular case of a finite-memory self-exciting process is the class of *Renewal processes*. A self-exciting point process with independent, identically distributed durations is termed an *ordinary Renewal process*. If all durations but the first (which is measured from the initial moment t_0) are identically distributed then the process is called a *modified Renewal process.* In both cases we have an example of a 1-Memory Self-Exciting PP. The only memory of the process is the last occurrence time (in particular there is no memory of the total number of past occurrences). In intuitive terms one can think that there is an underlying clock which is set to zero at every event-point. This means that the conditional intensity function for these processes depends only on the (left-continuous version of the) backward recurrence time

$$\bar{U}_1(t) = t - T_{\bar{N}(t)}, \tag{14}$$

The basic idea behind Self-Exciting counting (and point) processes, that is, allowing the conditional intensity function to depend on the internal filtration of the counting process, can be taken a step further. By allowing the conditional intensity function to be a stochastic process adapted to the filtration generated by some (eventually multivariate) 'information' process X we arrive at the class of *Doubly Stochastic Poisson processes* also known as *Cox processes*.

Definition 12 ***(Doubly Stochastic Poisson Processes):***

Let $\{N(t); t \geqslant t_0\}$ be a counting process associated with a given point process, let $\{X(t); t \geqslant t_0\}$ be some left-continuous (multivariate) stochastic process, we say that $N(t)$ is a Doubly Stochastic Poisson counting process with intensity process $\{\lambda(X(t)); t \geqslant t_0\}$ *if for almost every* given *path $\{X(t,\omega); t \geqslant t_0, \omega \in \Omega\}$ of the process X, $N(t)$ is a Poisson counting process with conditional intensity* function $\lambda(X(t,\omega))$.

In basic terms this means that the intensity process for this class of Point Processes is a *deterministic* function of the stochastic process $\{X(t); t \geqslant t_0\}$.

2.1.3 The Likelihood functional for Doubly Stochastic Poisson Processes

If the 'information' process $\{X(t); t \geqslant t_0\}$ is observable, then we can use the result presented in equation (15) in Subsection 2.1.1 to obtain the Likelihood associated with a sample path from one of these PP. The main problem associated with ML estimation and inference for Doubly Stochastic Poisson Processes comes from the fact that the 'full information' process X may not necessarily be observable. That is, the filtration generated by the 'information' process X may be decomposable as $\mathcal{F}_t = \mathcal{F}_t^o \cup \mathcal{F}_t^*$, where the observable filtration $\mathcal{F}_t^o$ corresponds to the observable components of X, while $\mathcal{F}_t^*$ is a collection of subsets of Ω connected with the history of the remaining (unobservable) factors.

$$L(\lambda \mid \mathcal{F}_T) = \mathrm{E}\left[\exp\left\{\int_{t_0}^{T} \ln \lambda(X(\tau))\mathrm{d}N(\tau) - \int_{t_0}^{T} \lambda(X(\tau))\mathrm{d}\tau\right\}\middle| \mathcal{F}_T^o\right] \tag{15}$$

The multivariate integral implied by (15) represents a challenge for the practical implementation of estimation and inference procedures for this class of PP (when some of the components of X are *latent*). Monteiro (2008) presents a simulation study comparing three different solutions for solving this problem. These three different methods for (numerically) evaluating the expectation appearing in (15), are applied to a new class of *generalized point processes* introduced in Koopman et al. (2008).

Next I introduce a theorem expressing a limit to the identification of latent intensity factors in multiplicative intensity point process models.

Theorem 5 ***(The Identifiability of Latent Intensity Factors)***. *Let $N(t)$ be an $\mathcal{F}_t$-adapted counting process observed over $[t_0, \infty]$, having an absolutely continuous, unbounded, $\mathcal{F}_t$-compensator process $\Lambda(t)$. Let the corresponding sequence of occurrence times be denoted as $\{T_n\}$. Assume that the corresponding intensity process can be factored as $\lambda(t) = \phi(t)\psi(t)$, where $\phi(t)$ is a non-negative, $\mathcal{F}_t^o$-adapted, càglàd process, while $\psi(t)$ is a non-negative $\mathcal{F}_t$-adapted, càglàd process ($\sigma(N(t)) \subseteq \mathcal{F}_t^o \subseteq \mathcal{F}_t$), define $\Phi(t) = \int_{t_0}^{t} \phi(\tau)d\tau$. Further assume that $\Phi(\infty) = \lim_{t\to\infty} \Phi(t) = \infty$.*

Then, from any given realization of the point process $N(t)$ over a finite time-window $[t_0, T]$, it is not possible to conduct statistical inference regarding the individual *values $\psi(t)$ with $t \in (T_{n-1}, T_n \wedge T)$ and $n = 1, \ldots, \bar{N}(T) + 1$).*

Proof. We start by applying Theorem 4, from where we are able to conclude that the time-changed counting process $\tilde{N}(u) = N(\Phi^{-1}(u))$ has the $\tilde{\mathcal{F}}_u$-predictable intensity process $\tilde{\lambda}(u) = \psi(\Phi^{-1}(u)) = \psi(t)$. This means that the $\tilde{\mathcal{F}}_u$-likelihood functional can be written as

$$L\left(\tilde{\lambda}\middle|\, \tilde{\mathcal{F}}_{\tilde{T}}\right) = \left(\prod_{n=1}^{\tilde{N}(\tilde{T})} \psi(\Phi^{-1}(u_n))\right) \exp\left[-\sum_{n=1}^{\tilde{\tilde{N}}(\tilde{T})+1} \int_{u_{n-1}}^{u_n \wedge \tilde{T}} \psi(\Phi^{-1}(u))\mathrm{d}u\right], \tag{16}$$

where $u_n = \Phi(T_n)$, $\tilde{T} = \Phi(T)$. Obviously, $L\left(\tilde{\lambda}\middle|\tilde{\mathcal{F}}_{\tilde{T}}\right) = L\left(\lambda \mid \mathcal{F}_T\right)$. Consider the process ψ^* constructed in the following way.

$$\psi^*(t) = \begin{cases} \psi(T_n) & \text{for } t \in \left(\Phi^{-1}(\frac{u_{n-1}+u_n}{2}), T_n\right] \\ 2\left[\int_{u_{n-1}}^{u_n \wedge \tilde{T}} \psi(\Phi^{-1}(u))\mathrm{d}u\right]/(u_n - u_{n-1}) - \psi(T_n) & \text{for } t \in (T_{n-1}, \Phi^{-1}(\frac{u_{n-1}+u_n}{2})] \end{cases} \tag{17}$$

The process ψ^* has sample paths which are càglàd piecewise constant. Statistically, this process is completely indistinguishable from the *true* latent process ψ. In fact for any realization of the $N(t)$ process over the interval $[t_0, T]$ both processes ψ and ψ^* have the same likelihood functional (16). However, process ψ^* only requires the estimation of two unknown constants for each spell $(T_{n-1}, T_n \wedge T)$. ■

This theorem establishes the limits to the identifiability of latent intensity factors from point process data. Although nothing can be inferred with regard to isolated values $\psi(t)$ when $t \in (T_{n-1}, T_n \wedge T)$ and $n = 1, \ldots, \bar{N}(T) + 1)$, from (16) it is clear that point process data *are* informative with regard to the integral $\int_{u_{n-1}}^{u_n \wedge \tilde{T}} \psi(\Phi^{-1}(u))\mathrm{d}u$. In most of what we have seen so far,

when we speak of a (orderly) univariate Random Point Process, in intuitive terms, we speak of a series of similar "events occurring in a one-dimensional continuum, usually time, the events being distinguishable only by where they occur, i.e., having no qualitative or quantitative information attached to them "(Lewis, 1972, page 14). However, in many situations these event-points correspond to *transitions* of some individual or observational unit between two well-defined *states* within some denumerable *state space*. This motivates the definition of a *Transition process*.

Definition 13 ***(Generalized Point Processes)***

A Transition Process*, also known as a* Generalized Point Process*, is a continuous-time stochastic process* $\{S(t) : t \in \mathbb{R}\}$ *where each* $S(t)$*, for a fixed* t*, is a discrete random variable taking values over some denumerable (fixed) state space.*

A particularly important class of continuous-time discrete-support stochastic processes corresponds to those processes that actually have both finite-support and finite memory.

2.2 Continuous-Time Markov Chains

This section reviews some fundamental concepts and results from the theory of continuous-time finite-state Markov Processes, also known as continuous-time *Markov Chains*. These elements provide a useful background to the class of *reduced form* credit risk models (see for example Jarrow et al. 1997). A comprehensive introduction to Markov theory is provided in Grimmet and Stirzaker (1992). Isaacson and Madsen (1976) is a detailed treatment of Markov chains with an emphasis on engineering and reliability applications.

Consider a continuous-time discrete-valued stochastic process defined on the interval $[t_0, \infty)$, $\{S(t); t \geqslant t_0\}$ and assume that the state space $\mathbb{S}$ (in which $S(t)$ takes its values) is finite i.e. $\mathbb{S} = \{1, \ldots, s\}$. In order to avoid some serious technical difficulties that arise from the 'point wise' definition of $S(t)$, I will assume for the remainder of this section that this stochastic process obeys the càdlàg assumption. This means that *with probability one the trajectories of* $S(t)$ *are Right-Continuous with Left-Limit functions of time.* This type of trajectories admits a countable representation,

$$\{(S_n, T_n), n \in \mathbb{N}\},$$

where the $T_n, n \in \mathbb{N}$ sequence consists of the *transition (or occurrence) times*, and $S_n = S(T_n)$.

Definition 14 ***(Markov Chain):***

The continuous-time finite state-space stochastic process $\{S(t); t \geqslant t_0\}$ *is termed a (1*st *order) Continuous-Time Markov Chain (CTMC) if it satisfies the Markov property:*

$$P[S_n = j \mid S_{n-1} = s_{n-1}, \ldots, S_0 = s_0] = P[S_n = j \mid S_{n-1} = s_{n-1}], \tag{18}$$

for all natural numbers $j, s_0, ..., s_{n-1} \leqslant s$ *and any arbitrary sequence* $t_0 < t_1 < \ldots < t_{n-1} < t_n$ *of transition times.*[6]

This means that the evolution of the chain in a finite time interval $[t', t]$ does not depend on the history before time t', i.e. it is a memoryless process (or more exactly there is only 'memory' of the present state).

This makes it possible to define a matrix function of t' and t alone, whose $(i, j)^{\text{th}}$ entry gives the probability that the chain will be in state j at time t given that it was in state i at time t'.

[6] A p^{th} order Markov Chain satisfies a generalized version of the Markov property, $\mathrm{P}[S_n = j \mid S_{n-1} = s_{n-1}, \ldots, S_0 = s_0] = \mathrm{P}[S_n = j \mid S_{n-1} = s_{n-1}, \ldots, S_{n-p} = s_{n-p}]$ and so memory of the past is limited to the last p-steps

Definition 15 ***(Transition Probabilities Matrix):***

This matrix determines the evolution of the Markov chain. Entry (i,j) *of this matrix is defined by*

$$p_{ij}(t,t') = P[S(t') = j \mid S(t) = i] \geqslant 0, \tag{19}$$

Note that each row of this matrix sums up to one, as it contains a discrete probability function.

Theorem 6 *The Family of matrices* $\{P(t,t'); t' \geqslant t \geqslant t_0\}$ *for Continuous-time Markov Chains satisfies the following conditions,*

1. $P(t,t) = I_s$,
2. $P(t,t')$ is a stochastic matrix (i.e. all the entries are non-negative and all rows add up to one, $P(t,t')\mathbf{1}_s = \mathbf{1}_s$),
3. $P(t,t') = \mathrm{P}(t,u)P(u,t'); t \leqslant u \leqslant t'$ (Chapman-Kolmogorov Equations).

For a proof of this result see Grimmet and Stirzaker (1992).

Definition 16 ***(Generator Matrix):***

This matrix is the equivalent, for continuous time Markov chains, to the 1-step transition matrix for discrete time Markov chains, and to the intensity function for a Poisson PP. It is defined as

$$G(t) = \frac{\partial}{\partial t'} P(t,t') \mid_{t'=t} \tag{20}$$

Under the càdlàg assumption we have that:

$$\lim_{t' \downarrow t} P(t,t') = I_s$$

i.e. the entries of the transition probability matrix $P(t',t)$ are continuous functions of the second argument, and thus, it follows that the sum of each row of $G(t)$ is identical to zero,

$$G(t)\mathbf{1}_s = \mathbf{0}_s \ \text{ (a column vector).} \tag{21}$$

Additionally, the diagonal elements of $G(t)$ are non-positive. In fact with $\lambda_{ii}(t)$ denoting one such component of $G(t)$ we have,

$$\lambda_{ii}(t) = \lim_{t' \downarrow t} \frac{p_{ii}(t,t') - 1}{t' - t},$$

and as $0 \leqslant p_{ii}(t,t') \leqslant 1$, we clearly obtain $\lambda_{ii}(t) \leqslant 0$ by the continuity of $p_{ii}(t,t')$.
On the other hand, for the non-diagonal elements of G we have,

$$\begin{aligned} \lambda_{ij}(t) &= \lim_{t' \downarrow t} \frac{p_{ij}(t,t')}{t'-t}, \\ \text{and so } \lambda_{ij}(t) &\geqslant 0 \text{ by the continuity of } p_{ij}(t,t'). \end{aligned}$$

So equation (21) together with these results means that

$$\lambda_{ii}(t) = -\sum_{j=1, j\neq i}^{s} \lambda_{ij}(t). \tag{22}$$

Theorem 7 *The Transition and Generator matrices are related by the following system of partial differential equations (the Kolmogorov 'Evolution' Equations),*

$$\frac{\partial P(t,t')}{\partial t'} = P(t,t')\,G(t'), \text{ (Forward equation)} \tag{23}$$

$$\frac{\partial P(t,t')}{\partial t} = -G(t)\,P(t,t'). \text{ (Backward equation)} \tag{24}$$

Theorem 8 *The Transition Probabilities matrix is recoverable from the Generator matrix by finding the unique solution to the Kolmogorov Equations that also satisfies the auxiliary condition $P(t,t) = I_s$. This solution is explicitly given by*

$$P(t,t') = \prod_{t}^{t'} (I_s + G(\tau)\,d\tau), \tag{25}$$

where $\prod_a^b (I + M(\tau)d\tau)$, for a given (matrix) function M, denotes the product integral of M.

The product integral $\prod_a^b (I + M(\tau)\mathrm{d}\tau)$ represents a generalization of the usual (finite) product, just like the usual integral generalizes the summation operator. The next result is sometimes useful for the numerical evaluation of the product integral. For an exposition on product integration, and a more detailed treatment of the results presented here, see Gill (2001), Andersen et al. (1993, Section II.6) and Goodman and Johansen (1973).

Theorem 9 *The product integral $\prod_a^b (I + M(\tau)d\tau)$ can be expressed as the following infinite Péano series*

$$\prod_{a}^{b} (I + M(\tau)\,d\tau) = I + \sum_{p=1}^{\infty} \int_{a \leqslant \tau_1 < \ldots < \tau_p \leqslant b} \ldots \int M\,d\tau_1 \ldots M\,d\tau_p. \tag{26}$$

An important particular case of a continuous-time Markov chain arises when the *transition intensities*, that is, the individual entries of the Generator matrix, are constant.

Definition 17 ***(Homogeneous chains):***

A Markov Chain $\{S(t); t \geqslant t_0\}$ *with transition probabilities matrix* $P(t, t')$ *is called* homogeneous *iff*

$$P(t, t') = P(t_0, t_0 + (t' - t)), \forall t' \geqslant t \geqslant t_0. \tag{27}$$

This means that the transition probabilities matrix does not depend on historical time, but only on the amount of time elapsed between the initial and final moments. In such case, we can define for every displacement $\tau \geqslant 0$, $P_\tau = P(t_0, t_0 + \tau)$.

2.2.1 Generator and Transition matrices for Homogeneous Chains

For homogeneous continuous-time Markov-Chains the Transition Probabilities matrix, (19), as seen, is function only of the displacement τ between the initial and final time points, accordingly, if $\lim_{\tau \downarrow 0} P_\tau = I_s$, the Generator matrix (20) is now a constant matrix

$$G = \frac{\mathrm{d}}{\mathrm{d}\tau} P_\tau \mid_{\tau=0} . \tag{28}$$

In this case the Kolmogorov Equations assume a particularly simple form.

Theorem 10 *For a Homogeneous Continuous-time Markov Chain where the Transition matrix is a continuous function of* τ *the Generator and Transition matrices are related by the following ordinary differential equations (Kolmogorov equations),*

$$\frac{\mathrm{d}}{\mathrm{d}\tau} P_\tau = P_\tau G = G P_\tau. \tag{29}$$

Theorem 11 *The Transition Probabilities matrix is recoverable from the Generator matrix by finding the unique solution to the Kolmogorov DEs that satisfies the auxiliary condition* $P_0 = I_s$. *This solution, known as the matrix Exponential function, is explicitly given by:*

$$\begin{aligned} P_\tau &= \exp[\tau G] \ \textit{or equivalently} \\ P_\tau &= \sum_{n=0}^{+\infty} \frac{\tau^n}{n!} G^n. \end{aligned}$$

2.3 Continuous-time Semi-Markov Processes

Continuous-time Finite State-Space Markov processes have a very limited amount of memory. Only the current state is recorded at any given moment. This may be too restrictive to model many real-life problems.

A slightly larger class of continuous-time finite state-space processes, which include the Markov chains as a special case, is the class of continuous-time semi-Markov processes. Semi-Markov processes keep track not only of the present state, but also of the elapsed duration (or sojourn time) in that state. For this class of processes equation (18) is replaced by

$$\mathrm{P}\left[S_{n+1}=j, X_{n+1}\leq\tau \mid (S_0,T_0),\ldots,(S_n,T_n)=(i,t)\right]=Q_{ij}\left(t,\tau\right), \tag{30}$$

Where Q_{ij} is a nondecreasing measurable function of the second argument (called a *subdistribution function*) and satisfies

$$\sum_{j=1}^{s}Q_{ij}\left(t,\infty\right)=\lim_{\tau\to\infty}\sum_{j=1}^{s}Q_{ij}\left(t,\tau\right)=1, \tag{31}$$

for every $t\in\mathbb{R}_0^+$. In intuitive terms, these processes are characterized by the joint distribution $Q_{ij}\left(t,\tau\right)$ of the destination state j and holding time $X_n=T_n-T_{n-1}$ on the current state i, given that both this and the precise moment t when the system entered it are known. The matrix-valued function $Q\left(t,\tau\right)$ is called the *semi-Markov Kernel*.

This means that for a Continuous-time semi-Markov Chain the entries of the generator matrix are stochastic processes (instead of deterministic functions of time) which depend deterministically on the backward-recurrence time. For this reason semi-Markov processes are also known as Markov Renewal processes. Define for every $t\in\mathbb{R}_0^+$

$$H_i\left(t,\tau\right)=\sum_{j=1}^{s}Q_{ij}\left(t,\tau\right). \tag{32}$$

Due to condition (31), $H_i\left(t,\tau\right)$ is a d.f. on $\bar{\mathbb{R}}_0^+$ with respect to the second argument τ. This can be interpreted as the d.f. of the sojourn times in state i that start at time t. If $Q_{ij}\left(t,\tau\right)$, as a function of the second argument τ, is absolutely continuous with respect to Lebesgue's measure, then the partial derivative of $Q_{ij}\left(t,\tau\right)$ with respect to τ is well defined,

$$q_{ij}\left(t,\tau\right)=\frac{\partial Q_{ij}\left(t,\tau\right)}{\partial\tau}.$$

2.3.1 Semi-Markov Transition Probabilities

In the current context, and due to the influence of the elapsed duration over the semi-Markov kernel, the direct counterpart to the Markov transition probabilities (19) are the following 'Renewal' transition probabilities

$$p_{ij}\left(t,\tau\right)=\mathrm{P}\left[S\left(t+\tau\right)=j\mid N\left(t\right)-\bar{N}\left(t\right)=1,S_{N(t)}=i\right],\tau>0, \tag{33}$$

these are the probabilities of finding the unit in state j, τ units of time after it entered state i, given that this transition happened at (chronological) time t.
These probabilities satisfy (Pyke, 1961) the following system of non-homogeneous Volterra integral equations of the second kind on two independent variables (t and τ)

$$p_{ij}(t,\tau)=\delta_{ij}(1-H_i(t,\tau))+\sum_{k=1}^{s}\int_0^\tau p_{kj}(t+u,\tau-u)q_{ik}(t,u)\mathrm{d}u,\quad i,j=1,\ldots,s, \tag{34}$$

where δ_{ij} denotes Kronecker's symbol. These equations are the direct counterpart to the Kolmogorov 'evolution' equations in the Markov setting. Monteiro et al. (2006) gives, for the first time, a formal proof that equations (34) yield a unique, Fréchet differentiable, solution in terms of the *subdensity functions* q_{ij}. Monteiro et al. (2006) further introduces, for the first time, a feasible estimator for the matrix of semi-Markov transition probabilities based on window-censored event-history data.

3 Observation Driven Point Process Models

This section describes the main specifications proposed in the financial Econometrics literature for dealing with point processes that evolve with after-effects. The focus in this section is on observation driven models. For all the different specifications in this section, it is assumed that all relevant information is available to the Statistician, i.e. that the relevant filtration is fully observable.
There are two main approaches for modeling a point process evolving with after-effects.

1. Specifying directly the *forward-occurrence density* of a conditionally orderly process, that is, the probability density function of the next inter-arrival time (or *duration*), conditional upon some particular filtration. This approach is somewhat limited. First, because the

conditioning information set is only updated at each occurrence time, it is not possible to assess the impact of covariates varying between event-times. As a consequence, and strictly speaking, 'duration models' can only be applied to univariate point processes. This is due to the fact that the occurrence of events of a different type during a duration cannot be included in the internal filtration of the model. This problem can be circumvented by using a *marked* 'duration model' for the pooled counting process $N(t) = \sum_{s=1}^{S} N_s(t)$ associated with a $S-$dimensional PP. In this context, the mark M_n would correspond to the specific event type (i.e. $\mathbf{M} = 1, \ldots, S$). However, not many authors seem to have taken this path (see also Bowsher, 2007).

2. Specifying directly the Conditional Intensity of the point process. This is a far more flexible approach. The filtration is updated continuously, allowing both the inclusion of general time-varying covariates and the occurrence of events of several different types, making it possible to model multivariate PPs.

In principle, it would also be possible to parameterize directly the survivor function, but apparently, not so many authors have followed this approach, which would suffer from the same limitations as the one resulting from the use of the forward occurrence density.

3.1 The ACD class of models

This class of processes, introduced by Engle and Russell (1998), is a good example of the first approach mentioned above. The basic underlying idea consist in specifying directly the next duration as the product of a parametric (scaling) function of past durations by an i.i.d. noise process with positive support. It is in fact a particular case of the Multiplicative Error Model introduced in Engle (2002). Therefore, this model has the same general structure as the GARCH model. It is, in fact, an example of a self-exciting point process model where the conditional intensity function explicitly depends on the backward-recurrence time (14).

3.1.1 Econometric specification

Let $\{N(t); T_1, T_2, \ldots, T_{N(t)}\}$ denote the past history of a (univariate) conditionally orderly point process over the interval $[t_0, t]$, the durations are given by $\tau_n = T_n - T_{n-1}$. Let ψ_n be the

conditional expectation of τ_n given the past sequence $\tau_1, \ldots, \tau_{n-1}$ of inter-arrival times. That is,

$$\psi_n \equiv \mathrm{E}\left[\tau_n \mid \tau_{n-1}, \ldots, \tau_1\right] = \psi_n\left(\tau_{n-1}, \ldots, \tau_1; \theta_\psi\right), \tag{35}$$

the ACD class of models consists of parameterizations of (35) jointly with the assumption

$$\tau_n = \psi_n \varepsilon_n. \tag{36}$$

The innovations process is assumed to be

$$\{\varepsilon_n\} \sim \text{i.i.d. with density } \mathrm{p}\left(\varepsilon; \theta_\varepsilon\right). \tag{37}$$

Additionally, ε has unit mean, and θ_ψ and θ_ε are fixed unknown vectors of parameters.
Clearly, this definition allows a broad class of point process models. By using different functional forms in (35) and allowing different distributions for the multiplicative error term (37) we can arrive at different particular ACD models.
The conditional intensity function implied by this mechanism can be deduced as follows. Let

$$\lambda_0\left(\varepsilon\right) = -\frac{p\left(\varepsilon; \theta_\varepsilon\right)}{\int_\varepsilon^\infty p(\tau; \theta_\varepsilon)\mathrm{d}\tau}, \tag{38}$$

denote the hazard function corresponding to the density (37). Note that, as the durations are obtained as modulated versions of i.i.d. random draws from this density, the corresponding hazard function can be thought of as a *baseline hazard* (in the spirit of the proportional hazards model of Cox, 1972, in particular, the expected value of the standardized durations ε_n is 1). Now, consider a *deformation of the time-axis*, such that in the new, transformed, time-axis we read the standardized durations

$$\varepsilon_n = \frac{\tau_n}{\psi_n}.$$

The image of the original PP over this new, transformed, time-axis constitutes a Renewal process, due to the i.i.d nature of the standardized durations. Accordingly, the corresponding conditional intensity process is

$$\lambda^*\left(t^*\right) = \lambda_0\left(t^* - t^*_{\bar{N}(t^*)}\right), \tag{39}$$

where t^* denotes time measured over the transformed time-axis. This implies the following intensity over the original time-axis

$$\lambda\left(t\right) = \lambda_0\left(\frac{t - t_{\bar{N}(t)}}{\psi_{\bar{N}(t)+1}}\right)\frac{1}{\psi_{\bar{N}(t)+1}}. \tag{40}$$

The previous argument also shows that we can think of the ACD model as an *Accelerated Failure Time (AFT) model.* The past of the process changes the speed at which time will flow over the next duration (through the scaling effect that ψ_n imposes to the 'natural' durations ε_n). In contrast, in classic AFT models this change in 'speed' is driven by exogenous covariates. In computational terms, this specification raises two problems. First, the admissible class of parameterizations of (35) is restricted to either strictly positive functions of $(\tau_{n-1}, \ldots, \tau_1; \theta_\psi)$ (for any θ_ψ), or to values of θ_ψ that keep $\psi_n(\tau_{n-1}, \ldots, \tau_1; \theta_\psi)$ strictly positive for all possible durations $\tau_{n-1}, \ldots, \tau_1$. In the later case, numerically maximizing the implied likelihood may be problematic. Second, $p(\varepsilon; \theta_\varepsilon)$ is restricted to densities with strictly positive support.

The simplest ACD model possible corresponds to a homogeneous Poisson Point Process with intensity

$$\lambda(t) = \frac{1}{\psi},$$

by setting (35) identical to a constant, and using a standard Exponential density for (37).

In most empirical applications of this model, (35) is simply taken to be a linear function of a finite number of past (observed) durations and past conditional expected durations. The so called ACD(p,q) model consists of (36) jointly with the following linear parameterization of (35),

$$\psi_n = \omega + \alpha(B)\tau_{n-1} + \beta(B)\psi_{n-1}, \tag{41}$$

where B denotes the usual lag operator, and

$$\begin{aligned} \alpha(B) &= \alpha_1 + \alpha_2 B + \ldots + \alpha_{p-1} B^{p-1}, \\ \beta(B) &= \beta_1 + \beta_2 B + \ldots + \beta_{q-1} B^{q-1}. \end{aligned}$$

Note that this general ACD(p,q) model has 'full memory', in contrast to a finite memory PP, due to the autoregressive polynomial $\beta(B)$. This can be seen by noticing the similarity of (41) with the equation defining an Autoregressive Moving Average (ARMA) process. However, this similarity also shows that the impact of one specific duration, under suitable location of the complex roots of the $\beta(B)$ polynomial, will fade out exponentially. Therefore, specification (41) is a 'short-memory' duration model. This similarity also implies that we can obtain a finite p-memory Self-Exciting PP by taking the $\beta(B)$ polynomial identical to zero.

A convenient property of the (linear) ACD(p,q) model is that it allows a straightforward analytical computation of several unconditional moments of the generated duration sequences, by taking expectations on both sides of (41).

In the original article by Engle and Russell (1998), most of the emphasis was placed on two

particular choices for the innovations process (37). First of these, the EACD(p,q) model is obtained by using an Exponential distribution for (37) jointly with the linear specification (41) for the conditional duration (35). A particularly simple form is the EACD(1,1)

$$\psi_n = \omega + \alpha\tau_{n-1} + \beta\psi_{n-1}. \tag{42}$$

In this model the unconditional mean duration is

$$\mu = \frac{\omega}{1-\alpha-\beta},$$

and the conditional and unconditional variances of the durations are given by

$$\begin{aligned} V\left[\tau_n \mid \tau_{n-1}, \ldots, \tau_1\right] &= \psi_i^2, \\ \mu^2\left(\frac{1-\beta^2-2\alpha\beta}{1-\beta^2-2\alpha\beta-2\alpha^2}\right) &= \sigma^2. \end{aligned}$$

Accordingly, this model will exhibit excess dispersion in the corresponding unconditional distribution of the generated durations whenever $\alpha > 0$, this is a feature often observed in duration data sets.

The alternative particular parameterization, the WACD(p,q) model, uses a Weibull distribution for the innovations process (37) together with the linear process (41). The hazard function associated with a Weibull distribution, with parameters γ and κ, is given by

$$\lambda(\tau) = \gamma\kappa^{\gamma}\tau^{\gamma-1}. \tag{43}$$

Other possible distributions for the innovations are the Generalized Gamma, Log-normal and Log-Logistic distributions (see for example Kalbfleisch and Prentice, 2002, section 2).

A further extension of the basic ACD specification consists in including the values of marks associated with the previous m event-points in the mean equation (41). That is,

$$\psi_n = \omega + \alpha(B)t_{n-1} + \beta(B)\psi_{n-1} + \gamma'(B)\,\mathbf{z}_{n-1}, \tag{44}$$

with $\gamma(B) = \gamma_1 + \gamma_2 B + \ldots + \gamma_{m-1}B^{m-1}$ and $\mathbf{z}$ denotes a vector of marks. Again the need to insure a positive ψ_n requires the use of (non-trivial) constraints over the parameters in (44).

It is equally possible to include deterministic calendar effects, like 'time-of-day' effects for transaction data. This can be achieved by assuming that the conditional expected duration is affected by a deterministic function of the corresponding starting moment. For example, using the multiplicative form

$$\mathrm{E}\left[\tau_n \mid \tau_{n-1}, \ldots, \tau_1\right] = \varphi(T_{n-1};\theta_\varphi)\psi_n\left(\tilde{\tau}_{n-1}, \ldots, \tilde{\tau}_1;\theta_\psi\right),$$

where $\tilde{\tau}_n = \tau_n/\varphi(T_{n-1};\theta_\varphi)$ corresponds to the "diurnally adjusted" durations.

A plethora of different extensions and variations on the basic ACD(p,q) model have been

proposed in the literature. Next, I provide a brief description of the main ideas behind some of these specifications.

1. The class of *Log-ACD models*, introduced by Bauwens and Giot (2000) circumvents one of the problems connected with the linear ACD(1,1) model, namely the need for imposing constraints over the model parameters to insure the positivity of ψ_n, by resorting to the logarithmic transformation. That is, equation (42) is replaced by

$$\ln \psi_n = \omega + \alpha \ln \tau_{n-1} + \beta \ln \psi_{n-1}. \tag{45}$$

2. Generalizing the class of Log-ACD models, Fernandes and Grammig (2006) proposed the *Augmented ACD (AACD)* by using a Box-Cox transformation. Therefore (45) is replaced by

$$\frac{\psi_n^\delta - 1}{\delta} = \omega + \alpha \psi_{n-1}^\delta \left[|\varepsilon_{n-1} - b| - c|\varepsilon_{n-1} - b|\right]^\nu + \beta \frac{\psi_{n-1}^\delta - 1}{\delta}, \tag{46}$$

where $\delta > 0$, $\nu > 0$, b and c are unknown parameters. The factor $\left[|\varepsilon_{n-1} - b| - c|\varepsilon_{n-1} - b|\right]^\nu$ appearing in (46) is the so-called "news impact function."

3. In order to circumvent the 'short-memory' characteristics associated with the linear conditional duration process (41), Koulikov (2002), introduces a class of long-memory positive weakly stationary random variables. Equation (42) is replaced by

$$\psi_n = \omega + \alpha(1 - \beta B)^{-1}(1 - B)^{-d}(\tau_{n-1} - \psi_{n-1}), \tag{47}$$

where $0 < d < 1$ and the negative fractional power of the back-differencing operator can be obtained from the expansion $(1 - B)^{-d} = 1 + \sum_{j=1}^{\infty} c_j B^j$. The coefficients c_j of this expansion can be obtained recursively from $c_j = c_{j-1} \frac{j-1+d}{j}$, starting from $c_1 = d$.

4. A *Threshold ACD (TACD)* model was proposed by Zhang et al. (2001). The main idea behind this specification, is to use an observable variable (for example the previous duration) to select one of P different regimes. Each regime has its own conditional mean equation (42) and error distribution (37). Regime j is chosen if (for example) $\tau_{n-1} \in [r_{n-1}, r_n)$, where $0 < r_0 < r_1 < \ldots < r_J = \infty$ are the threshold parameters. For fixed values of the threshold parameters the remaining parameters of the TACD model can be estimated by ML. Performing a grid-search over the threshold values is a feasible solution for obtaining the corresponding ML estimates.

5. Meitz and Teräsvirta (2006) introduce a class of *Smooth Transition ACD (STACD)* models. In contrast with the (discrete) regime switching TACD model, where the DGP 'jumps'

between several different regimes, durations in the STACD model result of a continuous 'mixture' of several different conditional distributions. A particular case of this specification is as follows. The conditional duration is given by

$$\psi_n = \omega + \beta\psi_{n-1} + \alpha\tau_{n-1} + (\omega' + \alpha')G(\ln \tau_{n-1}), \tag{48}$$

with ω' α' denoting additional parameters and $G(.)$ is the so-called 'transition function.' In general $G(.)$ can be any non-negative, bounded real function of one real variable. Usually $G(.)$ takes values between 0 and 1. A particular choice for the transition function is the logistic form

$$G(\ln \tau_{n-1}; \zeta, r_1, \ldots, r_J) = \left[1 + \exp\left(-\zeta \prod_{j=1}^{J} \ln \tau_{n-1} - r_j\right)\right]^{-1} \tag{49}$$

where $\zeta > 0$, $r_1 \leq \ldots \leq r_J$ are unknown parameters. The integer J is normally chosen *a priori* and determines the shape of G. The STACD model encompasses several particular cases of the TACD model. A further extension would consist in extending the scope of the transition function in order to include the lagged conditional duration term $\beta\psi_{n-1}$, yielding a *Time-Varying ACD (TVACD)* model.

6. Droost and Werker (2004) propose a *semi-parametric* class of ACD models by relaxing both the distributional and i.i.d. assumptions behind the innovations process (37). That is, while the conditional mean is specified as a parametric function (as for the other ACD models), the conditional distribution of the noise process ε_n is estimated directly from the data using a combination of kernel density and Nadaraya-Watson regression estimators.

3.1.2 Estimation and Inference

Observation driven ACD models belong to the class of Self-Exciting point processes. Therefore, the Likelihood associated with a particular ACD specification can be obtained by combining the general result (11) with the particular form of (40) implied by the chosen innovations density (37) and the precise parameterization (35) of ψ_n.

However, the construction of the precise data Likelihood implied by some of the more complex ACD specifications can be difficult. More important still, if the assumed distribution for the noise process is mis-specified, and unless this one belongs to the Exponential family (see Gouriéroux et al., 1984), the derived estimators will not be consistent. Engle and Russell (1998) show that the estimators of the ACD parameters derived from the Exponential noise

distribution satisfy the Quasi-ML(QML) properties. Therefore these estimators are consistent even under mis-specification of the noise distribution. Droost and Werker (2004) showed that the QML estimators based on the Gamma distribution are equally consistent but provide no efficiency gains over the ones derived from the Exponential distribution. This essentially means that, for obtaining the pseudo-Likelihood function implied by a particular ACD-type model, we can combine the result (11) with the particular conditional mean function (35) used, assuming the innovations to be i.i.d. Exponential. The estimators resulting from the maximization of this pseudo-Likelihood will be consistent.
Recently, Peiris et al. (2008) propose the estimation of both the Exponential and Weibull ACD models using the *Estimating Functions* approach of Godambe (1985).
Dynamic 'mixture' ACD models (for example, the STACD) on the contrary, cannot be estimated by ML due to the path dependence of conditional durations (a similar problem arises for Markov-Switching GARCH models, see Gray, 1996).

3.2 The ACI class of models

As mentioned in the introduction to this paper, while the original ACD specification of Engle and Russell (1998) models directly the forward-occurrence density (of a univariate point process) as a function of past observed durations, the Autoregressive Conditional Intensity model of Russell (1999) models directly the conditional intensity process (8). The added flexibility that results from the intensity approach makes it feasible dealing with multivariate point processes. Additionally, the ACI mechanism allows the inclusion of time-varying covariates in a regression framework. The basic idea of this model consists in using a Vector Autoregressive Moving Average (VARMA) process for describing the dependence of the (vector) conditional intensity on the past of the multivariate point process.

3.2.1 Econometric specification

Consider a set of S distinct (right-continuous) counting processes $N_s(t)$, with $s = 1, \ldots, S$, defined on a given filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, which are observed over the interval $[0, T]$. Assume that the *pooled* counting process $N(t) = \sum_{s=1}^{S} N_s(t)$ is *orderly* and the *compensator* $\Lambda_s(t)$ associated with $N_s(t)$ is absolutely continuous. The corresponding *intensity* process is denoted by $\lambda_s(t)$.

The ACI mechanism parameterizes each individual intensity component as

$$\lambda_s(t) = q_s(t)\exp(\beta_s' w(t) + \phi^s_{\bar{N}(t)+1})h_s(\bar{U}(t)), \tag{50}$$

where $q_s(t)$ is a deterministic function of (chronological) time, intended, for example, for modeling high-frequency ("intra-daily") seasonality effects. Typically, $q_s(t)$ is specified as a low-order spline function. The p dimensional vector $w(t)$ contains a collection of time-varying covariates relevant to all event types, while β_s stores the regression coefficients corresponding to the s^{th} point process. The multivariate baseline hazard function h_s introduces dependence on the vector $\bar{U}(t) = (\bar{U}^1(t), \ldots, \bar{U}^S(t))$. The quantities $\bar{U}^s(t) = t - t^s_{\bar{N}_s(t)}$ are the amount of time elapsed, i.e. the duration, since the last occurrence of an event of each type (known in the Point Processes literature as the *backward-recurrence time*). Therefore, it is possible to assess the impact over the s^{th} point process of the duration with respect to the last event, both of that same or of a different type. Note that, for each $s = 1, \ldots, S$ fixed, the point $t^s_{N_s(t)}$ satisfies $\Delta N_s(t) = 1$, and $\Delta N_s(\tau) = 0$ for every τ in the interval $\left(t^s_{N_s(t)}, t\right)$ whenever this is a non-degenerate one. This multivariate *baseline hazard* function h_s can, take different possible parametric forms. For example, the product of S Weibull or Burr hazards.

Although a general VARMA structure is possible, most commonly, $\phi_n = (\phi^1_n, \ldots, \phi^S_n)'$ is defined as a Vector Autoregressive (VAR) process of order one,

$$\phi_{n+1} = A\phi_n + B\iota_n\iota_n'\xi_n, \qquad \text{and} \qquad \phi_1 = \mathbf{0}, \tag{51}$$

where $\iota_n = (\Delta N_1(t_n), \ldots, \Delta N_S(t_n))'$ is a random *selection vector*, $A = (a_{rs})$ and $B = (b_{rs})$ are $(S \times S)$ matrices of unknown parameters. Note that the matrix A can be made dependent on ι_n, according to $A = [A_1 \ldots A_S]\,(\iota_n \otimes I_S)$, such that, when the last event was of type s, the $S \times S$ matrix A_s is used as the VAR(1) coefficient ($\otimes$ denotes the Kronecker product and I_S is the identity matrix of order S). This adds extra flexibility to the model by allowing a rich set of dependence patterns among the different components of the multivariate PP.

The vector $\xi_n = (\xi^1_n, \ldots, \xi^S_n)$ contains the innovations corresponding to each intensity process. Later in this section I shall discuss in detail two possible specifications for the noise process.

The A matrix determines the degree of persistence of each shock to the process ϕ.

When the last event was of the s^{th} type we have $\iota_n = \mathbf{e}_s$,where $\mathbf{e}_s$ denotes the s^{th} column of the identity matrix I_S, therefore, the s^{th} column of B is used for re-scaling the (univariate) innovation $\iota_n'\xi_n = \xi^s_n$. This means that the *instantaneous* impact of an event of type s over the intensity process of type r equals $b_{rs}\xi^s_n$, that is

$$\frac{\partial \phi^r_{n+1}}{\partial \xi^s_n} = b_{rs}.$$

The eigenvalues of the matrix of persistence parameters A are required to lie inside the unit circle in order for the process ϕ to be mean reverting to its unconditional mean of zero (see proposition 2 and ensuing discussion in Russell, 1999). The (vector-valued) noise process ξ_n can be defined in, essentially, two different ways, both resulting from a suitable application of the random time change theorem. In the original 1999 article by Russell the vector of innovations ξ_n is defined as

$$\xi_n = (1 - \Lambda_1(t^1_{N_1(t_n)-1}, t^1_{N_1(t_n)}), \ldots, 1 - \Lambda_S(t^S_{N_S(t_n)-1}, t^S_{N_S(t_n)}))', \tag{52}$$

where $\Lambda_s(a,b) = \int_a^b \lambda_s(t)\mathrm{d}t$. That is, the innovation associated with each point process, consists of the increment in the corresponding *compensator process* over the interval defined by the two most recent events of that type. Therefore, ξ^s_n is a Martingale difference sequence. When the last event was of the s^{th} type, only the s^{th} component of ξ_n is used for updating the vector ϕ. In this case, $\xi^s_n > 0$ (or $\xi^s_n < 0$) means that the model over predicted (respectively, under predicted) the time length between two consecutive events of type s. Direct application of the random time change theorem implies that the sequence of (scalar) innovations $\iota_n'\xi_n$ is an i.i.d. Exp(1) noise re-centered in zero.

The alternative solution for defining the innovations ξ_n was suggested in Bowsher (2007). Instead of using the increments in the compensated counting processes of each type, the idea consists in using the increments in the pooled (compensated) counting process. That is,

$$\xi_n = (1 - \Lambda_n)\mathbf{1}_S, \tag{53}$$

where $\mathbf{1}_S$ denotes an S column vector of ones and $\{\Lambda_n\}$ denotes the sequence of increments in the compensator associated with the pooled process,

$$\Lambda_n = \sum_{s=1}^{S} \Lambda_s(t_{n-1}, t_n), \quad n = 1, \ldots, N(T). \tag{54}$$

In this case $\iota_n'\xi_n > 0$ (< 0) means that the model over predicted (under predicted) the time length between any two consecutive events (i.e. regardless of their type). As previously mentioned, due to the random time change theorem, Λ_n is an i.i.d. Exp(1) noise while $\iota_n'\xi_n$ has zero mean. The computation of the ACI residuals, in both cases, is straightforward.

3.2.2 Maximum Likelihood estimation and Inference

As it was seen in Subsection 2.1, the Likelihood associated to a sample path from a general univariate Self-Exciting point process is readily available, see equation (11). For the present

case of a multivariate ACI process the key consists in applying the result (11) to the pooled counting process. This leads to

$$L\left(\lambda(\theta) \mid \mathcal{F}_T\right) = \prod_{s=1}^{S} \exp\left[\int_{t_0}^{t} \ln \lambda_s(\tau;\theta) \mathrm{d}N(\tau) - \int_{t_0}^{t} \lambda_s(\tau;\theta)\mathrm{d}\tau\right], \tag{55}$$

where θ denotes the full vector of parameters associated with a particular ACI specification. The computation of ML estimates has to be done using some numerical optimization algorithm, as, in general, there are no closed-form expressions obtainable from (55).

ML-based inference procedures are directly available. Additionally, due to the fact that under correct specification of the ACI model, the sequence of residuals ξ_n^s, with $s = 1, \ldots, S$ and $n = 1, \ldots, N_s(T)$ are i.i.d. unit Exponential distributed, we can use a Ljung-Box test over the autocorrelations of the estimated residuals as a diagnostic check on the adequacy of the estimated model. An additional 'goodness-of-fit' test is the excess-dispersion test introduced by Engle and Russell (1998).

3.3 Generalized Hawkes models

This intensity-based class of self-exciting multivariate point processes was proposed by Bowsher (2007), adapting and extending the seminal work of Hawkes (1971) to the analysis of financial data. Accordingly, Bowsher (2007) describes a simple data transformation for dealing with the fact that financial transaction data is not continuously recorded in time, due to the presence of the overnight period when financial markets are closed. The generalized Hawkes (g-Hawkes) model also allows the inclusion of exogenous variables in a regression framework that enables the assessment of their impact over the intensity of each univariate component of the multivariate PP. I start by describing the original, univariate, Hawkes (1971) model, this is then followed by the generalized counterpart introduced in Bowsher (2007) and then, finally, the full multivariate g-Hawkes model.

3.3.1 The Univariate Hawkes model

This is a case of a 'full-memory' self-exciting point process. The entire past history of the process over $[t_0, t)$ determines the present value of the conditional intensity process. Hawkes (1971) introduced the following particular parameterization of the intensity process

$$\lambda\left(t \mid N(t), T_1, T_2, \ldots, T_{N(t)}\right) = \omega + \int_0^t \pi\left(t-u\right) dN(u) \tag{56}$$

where t denotes calendar time, $T_1, T_2, \ldots, T_{N(t)}$ are the previous occurrence times and $\omega > 0$ is an unknown parameter. The integrand, $\pi\left(.\right)$, is a fixed function of time termed the *infectivity measure* in the classical literature due to the first applications of this model in epidemiology. The most common parameterization of π was first suggested by Hawkes (1971)

$$\pi\left(\tau\right) = \sum_{k=1}^{K} \alpha_k \exp(-\beta_k \tau), \tag{57}$$

where $\alpha_k \geq 0$ and $\beta_k \geq 0$ are model parameters. The order of the process K is either set a priori or chosen using model selection criteria.

This model was initially used in the field of seismology, see for example Ogata and Katsura (1986). The main feature of this model, besides its full memory, is the fact that the marginal impact of one event recorded at calendar time t_n is independent of the remaining history of the PP. The amount of time elapsed since t_n is the only factor determining the impact of this event over the current value of the conditional hazard rate; regardless of the number of events recorded between t_n and t. This property of the Hawkes model, as argued in Engle and Russell (1998), may render it inadequate for the purpose of analyzing financial transactions data. In fact some authors argue that financial markets evolve in *transaction* instead of *chronological time*.

In contrast with this view, Bowsher (2007) adapts the Hawkes specification for modeling financial transaction data. The g-Hawkes model makes use of a simple data transformation, designed for dealing with the overnight periods when no activity takes place on the financial markets. The transformed time axis is defined in the following way, the origin is set at 9:30 a.m. of the first recorded trading day. The overnight periods are then removed from the time axis, such that only the l working hours of each trading day (6.5 hours for most financial markets) are retained. This means that moment x (expressed in hours measured after 9:30 a.m.) in trading day d (an integer number) will appear as time-point $l \times (d-1) + x$ in the final data set. Formally, we have the following partition of the time axis

$$(0, +\infty) = (0, x_1] \cup (x_1, x_2] \cup \ldots \cup (x_{d-1}, x_d] \cup \ldots, \tag{58}$$

where $x_d = d \times l$ $(d = 0, 1, 2, \ldots)$. With this partition in place, Bowsher (2007) proposes a special Hawkes-type specification for the conditional intensity process.

3.3.2 The Univariate g-HawkesE(K) model

Consider a self-exciting, conditionally orderly point process defined over $[0, \infty)$ and equipped with the above-mentioned partition. Let the conditional intensity process of this PP follow the parameterization

$$\lambda\left(t \mid N(t), T_1, T_2, \ldots, T_{N(t)}\right) = \mu(t) + \sum_{k=1}^{K} \widetilde{\lambda}_k(t), \tag{59}$$

where $\mu(t)$ is a strictly positive deterministic function of time (designed to accommodate deterministic intra-daily patterns). Each stochastic component $\widetilde{\lambda}_k$ is obtained recursively. Starting from $\widetilde{\lambda}_k(0) = 0$, the recursion step is given by

$$\widetilde{\lambda}_k(t) = \pi_k \widetilde{\lambda}_k(x_{d-1}) \exp\left[-\rho_k (t - x_{d-1})\right] + \int_{[x_{d-1}, t)} \alpha_k \exp\left[-\beta_k (t - u)\right] \mathrm{d}N(u), \tag{60}$$

when $x_{d-1} < t \leqslant x_d$. Additionally, the following sign restrictions need to be imposed, $\pi_k \geqslant 0, \alpha_k \geqslant 0, \rho_k \geqslant 0$ and $\beta_k \geqslant 0$. Equations (59) and (60) mean that, added to the deterministic function $\mu(t)$, there are K stochastic components which account for both a *'spillover effect'* from the trading intensity of the previous trading day (obtained from the first term on the right-hand side of (60)), and the past trading intensity on that day (given by the second term). Both effects have an associated exponential *'rate of memory loss.'* Inside a specific trading day one individual event leads to a 'jump' of amplitude $\sum_{k=1}^{K} \alpha_k$ in the intensity λ. This increase will eventually fade out at exponential rate. More precisely, at the rate of $\exp\left[-t \times \left(\min_k \beta_k\right)\right]$. With regard to the 'spillover' term on the right-hand side in equation (60), it should be noted that the value of each stochastic component at the closing time on day $d-1$, included in $\lambda(t)$ with $x_{d-1} < t \leqslant x_d$, 'fades out' at the (distinct) rate $\exp\left[-t \times \left(\min_k \rho_k\right)\right]$.

The basic self-excitement mechanism of the g-HawkesE(K) model, built using weighted exponential response-functions, is easily extended to the multivariate case by including terms that account for possible *cross-effects* between the different components of the multivariate PP. Again, for simplicity of exposition and following Bowsher (2007) only the bivariate case is described. Constructing the general multivariate case is nonetheless straightforward. In particular, I introduce in this paper an innovative notation for describing the bivariate g-Hawkes that lends itself to a trivial generalization for handling the general multivariate case.

3.3.3 The Bivariate g-HawkesE(K) model

Consider two distinct, conditionally orderly, point processes with associated intensity processes $\lambda_1(t)$ and $\lambda_2(t)$ assembled into a (column) vector $\lambda(t) = (\lambda_1(t), \lambda_2(t))'$. The bivariate g-HawkesE(k) model assumes this vector intensity follows the parameterization:

$$\lambda(t) = \mu(t) + M(t)\,\mathbf{1}_K, \tag{61}$$

where $\mu(t) = (\mu_1(t), \mu_2(t))'$ is a two-dimensional deterministic function of time. The matrix $M(t) = (m_{ij}(t))$, $i, j = 1, 2$ has entry $m_{ij}(t) = \sum_{k=1}^{K} \overset{\sim(k)}{\lambda}_{ij}(t)$. Each individual stochastic component $\overset{\sim(k)}{\lambda}_{ij}(t)$ is defined similarly to the univariate case. That is, $\overset{\sim(k)}{\lambda}_{ij}(0) = 0$, and then

$$\overset{\sim(k)}{\lambda}_{ij}(t) = \pi_{ij}^{(k)} \overset{\sim(k)}{\lambda}_{ij}(x_{d-1}) \exp\left[-\rho_{ij}^{(k)}(t - x_{d-1})\right] + \int_{[x_{d-1},t)} \alpha_{ij}^{(k)} \exp\left[-\beta_{ij}^{(k)}(t-u)\right] \mathrm{d}N_j(u), \tag{62}$$

when $x_{d-1} < t \leqslant x_d$. The following restrictions are required, $\pi_{ij}^{(k)} \geqslant 0, \alpha_{ij}^{(k)} \geqslant 0, \rho_{ij}^{(k)} \geqslant 0$ and $\beta_{ij}^{(k)} \geqslant 0$. $N_j(u)$ stands for the counting process associated with the j^{th} component of the PP. The bivariate version basically adds the facility for cross-effects between the occurrence of events and the intensities of the different component PP.

3.3.4 Maximum Likelihood Estimation

The data Likelihood implied by this class of models follows directly from equation (11) and the specification of the conditional intensity process (61). This leads to a general Likelihood expression similar to (55). However, the structure of the transformation of the time-axis (58), allows writing the likelihood as a product of daily contributions. This decomposition in turn, allows the use of the recursive specification (62) in order to compute the log-likelihood in an efficient way. Additionally, the special parametric forms in (62) mean that the integral of the path of each s^{th} intensity component (appearing in (55)) can be computed analytically, see Bowsher (2007) for details.

4 Parameter Driven Point Process Models

In this section the focus is placed on point process models with unobserved components. The common feature across the different specifications in this section, is that the internal filtration generated by all these models can be decomposed as $\mathcal{F}_t = \mathcal{F}_t^o \cup \mathcal{F}_t^*$, where $\mathcal{F}_t^o$ corresponds to the observable information set, while $\mathcal{F}_t^*$ denotes the history of the unobserved components. The dynamic behavior of these models is driven not only by an observable filtration but also by latent components, thus combining aspects from both self-exciting and doubly stochastic point processes. The richer dynamic structure of this class of models provides added flexibility for describing the patterns in empirical point processes. This added flexibility, however, comes at a cost. ML estimation for this class of models is hindered by the need to integrate out the effect of the unobserved components. This is a common and well-known problem for parameter driven nonlinear or non-Gaussian dynamic statistical models. The data-density typically involves a high-dimensional integral, which has (due to the unavailability of exact closed-form solutions) to be evaluated either using simulation or other approximate methods.

4.1 Parameter driven ACD models

As seen in Subsection 3.1, the class of ACD models constitutes, by historical reasons, one of the main point process models used in the financial literature. Many different extensions and variations on the original specification of Engle and Russell (1998) have been proposed. Some of these were already mentioned in Subsection 3.1. Here I focus on those particular extensions of the ACD model that include latent components.

4.1.1 The Markov Switching ACD model

As mentioned in Subsection 3.1 ACD models are characterized by a particular shape for the conditional duration (35), the multiplicative error structure (36) and a particular set of stochastic assumptions for the noise process ε_n.

One particular extension of the basic ACD model mentioned in Subsection 3.1 was the TACD model of Zhang et al. (2001). The main characteristic of the TACD specification consists in combining different 'regimes,' i.e. the durations are generated according to several different conditional mean functions and innovation distributions. The particular regime used to generate the next duration is chosen according to the value of the previous observed duration. An alternative way to shift between several different regimes is a (hidden) *Markov switching* mechanism. That is, the particular regime (among J possible regimes) generating the next conditional

duration ψ_{n+1} is chosen according to the value of an unobserved discrete random variable r_{n+1} (with finite support $\mathbf{J} = \{1, \ldots, J\}$) following a (discrete) Markov chain. This is the main idea behind the *Markov Switching ACD (MSACD)* model introduced by Hujer et al. (2002).

The MSACD model is characterized by the multiplicative error structure (36) (implying $\mathrm{E}[\varepsilon_n] = 1$), the conditional mean depends directly on the unobserved regime variable r_n

$$\psi_{n+1} = \sum_{j=1}^{J} \mathrm{P}[r_{n+1} = j | \mathcal{F}_n; \theta] \psi_{n+1}^{(j)}, \tag{63}$$

where $\mathrm{P}[r_{n+1} = j | \mathcal{F}_n; \theta]$ is the probability that the next regime will be in state j, given the information set available at time T_n. The regime-specific conditional mean

$$\psi_{n+1}^{(j)} = \mathrm{E}[\tau_{n+1} | r_{n+1} = j, \mathcal{F}_n; \theta],$$

is normally specified according to an autoregressive specification of the form (41).

As mentioned the latent stochastic process $\{r_n\}$ follows a homogeneous (discrete) Markov chain, characterized by the 1-step transition matrix P, with entries $p_{ij} = \mathrm{P}[r_{n+1} = j | r_n = i]$. As with the conditional mean duration, the next conditional error distribution depends only on the current state r_n and the (observable) information set $\mathcal{F}_n$. That is, ε_{n+1} is drawn from $f(\varepsilon_{n+1} | r_{n+1} = j, \mathcal{F}_n; \theta)$.

There are two possible ways in which to specify the regime-specific conditional expected duration $\psi_{n+1}^{(j)}$. First, $\psi_{n+1}^{(j)}$ can be regressed only on previous expected durations and observed durations corresponding to that same regime j. Alternatively, $\psi_{n+1}^{(j)}$ can be written as an autoregressive function of both previous expected and observed durations regardless of the specific regime. This second possibility raises the problem of *path dependence*, that is, $\psi_{n+1}^{(j)}$ becomes dependent on the unobserved past trajectory $r_1, \ldots, r_n$. This requires a computationally expensive procedure, as all J^n possible past trajectories need to be considered. A simplifying aggregation procedure for this case was suggested by Gray (1996), in the context of Markov switching GARCH models (see also Hujer et al., 2002).

The data Likelihood for the MSACD class of models is given by the average conditional likelihood of the observable variables, taken over all possible trajectories $(r_1, \ldots, r_{N(T)})$ of the latent Markov chain. Evaluation of this Likelihood function is computationally demanding. Hujer et al. (2002) suggest a feasible procedure making use of the *Expectation-Maximization (EM) algorithm* of Dempster et al. (1977).

4.1.2 The Stochastic Conditional Duration model

The relationship between the Stochastic Conditional Duration (SCD) and the ACD models is similar to that between the Stochastic Volatility (SV) and GARCH models. Instead of specifying the expected conditional duration (35) as a function of previous observed durations, as in the ACD model, this expected duration is modeled as a latent stochastic process.
While the multiplicative error structure (36) is retained, the conditional expected duration equation (41) is replaced by

$$\ln \psi_n = \omega + \beta \ln \psi_{n-1} + \eta_n, \text{ with } |\beta| < 1, \tag{64}$$

where the latent noise process, $\eta_n \sim N(0, \sigma^2)$, is assumed independent of ε_n, given $\mathcal{F}_{n-1}$. The initial value of the latent conditional mean log-duration process, that is $\ln \psi_0$, is drawn from the 'steady state' distribution of $\ln \psi_n$. The SCD model implies a marginal distribution for the durations τ_n that results from mixing the (assumed) log-Normal distribution of ψ_n with the chosen distribution for ε_n. In general, it is not possible to compute these distributions (i.e. the unconditional one and the one conditional upon $\mathcal{F}_n$) analytically given a choice of a parametric family of distributions for ε_n. It is, however, possible to obtain these distributions by numerical integration (see Bauwens and Veredas, 2004).
In what follows I assume that the distribution of ε_n has finite moments of all orders. These moments are denoted by

$$g_p = \mathrm{E}\left[\varepsilon_i^p\right], \qquad p = 1, 2, \ldots$$

Two possible choices are the standard Weibull distribution $W(\gamma, 1)$ and the standard Gamma distribution $G(\nu, 1)$ for which

$$\begin{aligned} g_p &= \Gamma\left(1 + \frac{p}{\gamma}\right) \qquad \text{(Weibull)}, \\ g_p &= \frac{\Gamma(\nu + p)}{\Gamma(\nu)} \qquad \text{(Gamma)}. \end{aligned}$$

The sequence of durations τ_n constitutes a strictly stationary process under the restriction $|\beta| < 1$, which also implies the stationarity of the latent factor ψ. The unconditional moments of these processes are given by

$$
\begin{aligned}
\mu_\psi &= \exp\left[\frac{\omega}{1-\beta} + \frac{1}{2}\left(\frac{\sigma^2}{1-\beta^2}\right)\right], \\
\mu_\tau &= g_1\mu_\psi, \\
\sigma^2_\psi &= \mu^2_\Psi\left[\exp\left(\frac{\sigma^2}{1-\beta^2}\right) - 1\right], \\
\sigma^2_\tau &= \mu^2_t\left[\frac{g_2}{g_1^2}\exp\left(\frac{\sigma^2}{1-\beta^2}\right) - 1\right].
\end{aligned}
$$

See Bauwens and Veredas, (2004) for a proof of these results. The SCD model is able to generate a sequence of durations τ_n exhibiting excess dispersion if

$$\frac{\sigma^2}{1-\beta^2} > \ln\left(2\frac{g_1^2}{g_2}\right).$$

For Weibull distributed innovations, this condition holds if $\gamma \leq 1$ ($\nu \leq 1$ for the Gamma case) and $\sigma^2 > 0$ (even if $\beta = 0$).

The theoretical autocorrelation function (ACF) of the sequence of durations τ_n is given by

$$\rho_p = \frac{\exp\left(\frac{\sigma^2\beta^p}{1-\beta^2}\right) - 1}{\frac{g_2}{g_1^2}\exp\left(\frac{\sigma^2}{1-\beta^2}\right) - 1}.$$

This result implies that the ACF decreases geometrically with s. Therefore, the SCD model is, just like the ACD, a 'short-memory duration process. As seen on Section 2.1.3 Doubly

Stochastic Poisson Processes are not easy to estimate using Maximum Likelihood. Following Harvey et al. (1994) and Ruiz (1994), Bauwens and Veredas (2004) propose a Quasi-Maximum Likelihood (QML) approach obtained from the application of the Kalman filter to the state space representation of the SCD model

$$\ln\tau_n = \mu + \psi_n + \xi_n \qquad \text{(observation equation)}, \tag{65}$$

$$\ln\psi_n = \omega + \beta\ln\psi_{n-1} + \eta_n \qquad \text{(state equation)}, \tag{66}$$

where $\xi_n = \ln\varepsilon_n - \mu$, and $\mu = \mathrm{E}[\ln\varepsilon_n]$.

The Kalman filter would provide the exact Likelihood for this state space model (SSM) if the ξ_n disturbances were Normally distributed. Because this is not the case (unless the distribution of ε is taken to be log-Normal), maximizing the likelihood obtained from application of the Kalman filter to the SSM given by equations 65 and 66 is a pseudo ML and not a full ML procedure.

More recently, Bauwens and Galli (2007) describe the details of applying the Efficient Importance Sampling (EIS) algorithm of Richard and Zhang (2007) to the estimation of SCD models. As it would be expected, they find that (at the cost of an increased computational effort) the simulation-based EIS algorithm provides more accurate estimates of the model parameters, when compared with the original QML procedure suggested in Bauwens and Veredas (2004).

4.2 Stochastic Conditional Intensity processes

Introduced in Bauwens and Hautsch (2006a), this class of 'intensity based' point process models can be considered a parameter driven extension of the ACI model of Russell (1999). As it is the case with the SCD and ACD models, the SCI specification results from enlarging the filtration generated by the original ACI model with the σ-algebra generated by a single latent (univariate) stochastic process. That is, although the model can be applied to an arbitrary S-dimensional point process, it is assumed that a common latent factor influences (multiplicatively) the conditional intensity process of each individual component.

More formally, consider a set of S distinct (right-continuous) counting processes $N_s(t)$, with $s = 1, \ldots, S$, defined on a given filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, and observed over the interval $[0, T]$. Assuming that the *pooled* counting process $N(t) = \sum_{s=1}^{S} N_s(t)$ is *orderly* and the *compensator* $\Lambda_s(t)$ associated with $N_s(t)$ is absolutely continuous, there is a corresponding *intensity* process, denoted by $\lambda_s(t)$. In the broadest sense possible, the SCI class of models consists of all possible parametric specifications for the vector of intensities $(\lambda_s(t))$, where each component can be written multiplicatively as

$$\lambda_s(t) = \lambda_s^o(t) \exp(\sigma_s \psi(t)), \tag{67}$$

with $\psi(t)$ denoting a *univariate* unobserved stochastic process with piecewise-constant càglàd sample paths. In fact, the unobserved process cannot be identified between successive events of the pooled process (recall theorem 5). Only the *integral* of this latent process over a *complete spell* $(t_{n-1}, t_n]$ of the pooled process and its boundary values $\psi(t_n)$ are identifiable. Accordingly, it is reasonable to consider a process with piecewise-constant càglàd sample paths (only 'jumping' at the occurrence times t_n, with $n = 1, \ldots, N(T)$) as an approximation to the 'true' *dynamic unobserved heterogeneity* process. Clearly, the trajectories of such a process can be indexed by the left-continuous counting process $\bar{N}(t)$ associated with the pooled process. Finally, for this class of models, the *latent* information set $\mathcal{F}_t^*$ mentioned previously, corresponds to the history of the process $\psi(t)$. And therefore, it is updated only at the occurrence times of

the pooled counting process $N(t)$.

The observable intensity component $\lambda_s^o(t)$ can, in general, be any predictable process of the observable filtration $\mathcal{F}_t^o$. In contrast with $\mathcal{F}_t^*$, $\mathcal{F}_t^o$ can be updated *continuously*.

In practice, specific parametric processes have to be chosen for modeling both the latent and the observable components of λ_s. In Bauwens and Hautsch (2006a), the observable intensity factor λ_s^o is modeled as an ACI process, while an AR(1) specification driven by Gaussian innovations was chosen for the common latent process. That is

$$\psi_{\bar{N}(t)+1} = \rho\psi_{\bar{N}(t)} + \varepsilon_{\bar{N}(t)+1}, \qquad \varepsilon_{\bar{N}(t)+1} \sim \text{NIID}(0,1). \tag{68}$$

The latent innovations process is assumed to be independent of the series of increments in the compensator of the pooled process (54). This assumption is required to insure valid intensity components λ_s. The stationarity of the latent AR(1) process (obtained when $\mid \rho \mid < 1$) is a necessary condition for the stationarity of the SCI model.

Note that σ_s represents the conditional standard deviation of the s^{th} log-intensity component $\ln \psi_s$ given $\mathcal{F}$. For this reason, the latent innovations ε_n have unit variance (68).

The component-specific standard deviation σ_s means that a shock ε_n to the common unobserved factor will have a different impact over each individual component of the S-variate PP.

As mentioned above, the observable part of the intensity function for a SCI model λ_s^o, can be parameterized using the ACI mechanism, Bauwens and Hautsch (2006a) suggest the use of a product of Burr hazard functions for the baseline hazard function h_s discussed in Subsection 3.2. There is, however, one critical difference between the (original) ACI specification and the parameterization of the observable component of the intensity process in the SCI model, as suggested by Bauwens and Hautsch (2006a). If the innovations process ξ_n is computed following either (52) or (53) then, in both cases, ξ_n will be a function of previous values of the latent component. Recall that in the ACI model, the innovation associated with each individual counting process $N_s(t)$, at an occurrence time T_n of the pooled process $N(t)$, was either the increment in the s^{th} compensated counting process over the last complete spell of the associated point process, or the increment in the pooled compensated counting process. However, when an unobserved component ψ is added to the intensity processes, both Λ_n and $\Lambda_s(t^s_{N_s(t_n)-1}, t^s_{N_s(t_n)})$ become dependent on one or more past values of this latent factor. Clearly, this means that one cannot separate the computation of the ACI residuals from the problem of filtering the (latent) process ψ. In order to separate the computation of the innovations ξ_n from the filtering problem for ψ, Bauwens and Hautsch (2006a) suggest an alternative specification for the ACI innovations. The basic idea is to compute the ACI innovations process ξ based only on the observable intensity components λ_s^o. This simplifies considerably the computation of the (redefined) ACI residuals,

but at the cost of seriously hindering the determination of their exact distribution (which is still unknown). Under the approach of Russell (1999), Bauwens and Hautsch (2006a) suggest the following definition for the ACI residuals,

$$\xi_n = (-\gamma - \ln \Lambda_1^o(t^1_{N_1(t_n)-1}, t^1_{N_1(t_n)}), \ldots, -\gamma - \ln \Lambda_S^o(t^S_{N_S(t_n)-1}, t^S_{N_S(t_n)}))', \tag{69}$$

where $\gamma \approx 0,5772$ denotes the Euler-Mascheroni constant and $\Lambda_s^o(a,b) = \int_a^b \lambda_s^o(t)\mathrm{d}t$. The alternative specification for the noise process suggested by Bauwens and Hautsch (2006a), follows the approach suggested by Bowsher (2007)

$$\xi_n = (-\gamma - \ln \sum_{s=1}^{S} \Lambda_s^o(t_{n-1}, t_n))\mathbf{1}_S. \tag{70}$$

Clearly, in this last case, ξ_n does not depend on the type of the most recently observed event. The reason for taking the natural logarithm (apart from eventually leading to an increase in the numerical stability of the corresponding computations) is understandable mainly in the univariate case (i.e. $S = 1$). Only in this case it is possible to write the (univariate) disturbance term ξ_n as

$$\xi_n = -\gamma + \sigma_1 \psi_n - \ln \Lambda_1(t_{n-1}, t_n). \tag{71}$$

That is, because under correct model specification $\Lambda_1(t_{n-1}, t_n)$ is Exp(1) distributed, ξ_n is the sum of a re-centered standard Gumbel (minimum) random variable with the latent factor. However in the general (multivariate) case, the log transformation does not lead to a clear decomposition similar to (71). This is true both for the Russell and Bowsher specifications of the ACI innovations. Further note that equation (15) on page 458 of the article is clearly inconsistent.

4.2.1 Estimation and Inference for the SCI model

The main challenge in estimating this type of models comes from the presence of the latent factor, which must be 'integrated out' of the conditional (upon the true path of the latent process) likelihood function. The data Likelihood implied by the SCI model can be obtained by combining the results (15) and (55). That is, if the complete path of the latent process $\Psi_{\bar{N}(T)+1} = \{\psi_i\}_{i=1}^{\bar{N}(T)+1}$ would be known, then (55) would provide the data Likelihood for the SCI model when (67) is used. Let $L\left(\theta \mid \mathcal{F}_T^o, \Psi_{\bar{N}(T)+1}\right)$ denote this conditional Likelihood. Because $\Psi_{\bar{N}(T)+1}$ is unknown, the likelihood becomes

$$L(\theta \mid \mathcal{F}_T^o) = \int L\left(\theta \mid \mathcal{F}_T^o, \Psi_{\bar{N}(T)+1}\right) p(\Psi_{\bar{N}(T)+1})\mathrm{d}\Psi_{\bar{N}(T)+1}, \tag{72}$$

where and $p(\Psi_{\bar{N}(T)+1})$ denotes the (unconditional) density function of $\Psi_{\bar{N}(T)+1}$.
Several different approaches are available for evaluating the high-dimensional integral in (72). In Monteiro (2008), three different methods are applied to a multi-state (i.e. a generalized point process) extension of the SCI model introduced in Koopman et al. (2008). Bauwens and Hautsch (2006a) apply the simulation-based EIS algorithm of Richard and Zhang (2007).
Inference in parameter driven models is not only limited to the estimation and evaluation of hypothesis concerning the unknown parameters of the model. A central issue is the so-called *signal extraction* problem. This consists in two closely related problems, the *filtering* and the *smoothing* problems. Usually, the filtering problem consists in obtaining the conditional expectations $\mathrm{E}[\psi_{n+1}|\mathcal{F}^o_{t_n}]$, with $n = 1, \ldots, \bar{N}(T)$, of the latent stochastic process given past observations up to (and including) the previous moment t_n. The smoothing problem, on the other hand, consists in obtaining the conditional expectations $\mathrm{E}[\psi_n|\mathcal{F}^o_T]$, with $n = 1, \ldots, \bar{N}(T)+1$, of the latent stochastic process given all observations. The different feasible estimation methods that can be applied for evaluating the Likelihood (72) can also deal with these signal extraction problems.

5 Final remarks

This paper provided an introductory overview of the core financial econometrics literature dealing with models for randomly spaced data. The main concepts and results from the theory of (generalized) point processes were also briefly reviewed, in order to provide the adequate background for understanding the underpinnings of the models surveyed. A new result regarding the identifiability of latent intensity factors in point process models was introduced.
Statistical models for point processes, like other dynamic statistical models, can be classified as either *observation* or *parameter driven* models. Most of the econometric models for point processes, in the literature, are observation driven. This is perhaps due to the impact that the ACD model of Engle and Russell (1998) has had over the financial econometrics literature. Nevertheless, a fast-growing literature on point process models with *unobserved components* has recently appeared. This particular stream of literature started with the introduction by Hujer et al. (2002) of a parameter driven extension of the ACD model. The main obstacle to the wider use of some of these more flexible econometric models is, clearly, the complexity and computational load associated with their estimation. This issue, in particular for the class of generalized SCI point processes, was the focus of some of my previous work.

References

Andersen, P.K., Borgan, Ø., Gill, R.D. and N. Keiding (1993): *Statistical Models Based on Counting Processes,* Springer-Verlag, New York.

Aalen, O.O. and J.M. Hoem (1978): "Random Time Changes for Multivariate Counting Processes," *Scandinavian Actuarial Journal*, **2**, pp. 81-101.

Bauwens, L. and P. Giot (2000): "The logarithmic ACD model: An application to the Bid/Ask quote of two NYSE stocks," *Annales d'Economie et de Statistique*, **60**, pp. 117-149.

Bauwens, L., and N. Hautsch, (2006a): "Stochastic Conditional Intensity processes," *Journal of Financial Econometrics* **4**, pp. 450-493.

Bauwens, L., and N. Hautsch, (2006b): "Modelling Financial High Frequency Data using Point Processes," *Working paper Nr. 2006/80 - CORE.*

Bauwens, L. and D. Veredas (2004): "The stochastic conditional duration model: A latent factor model for the analysis of financial durations," *Journal of Econometrics*, **119**, pp. 381-412.

Bauwens, L. and F. Galli, (2009): "Efficient importance sampling for ML estimation of SCD models," *Computational Statistics and Data Analysis*, **53**(6), pp. 1974 - 1992.

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

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

Cox, D.R. (1972): "Regression models and life tables," *Journal of the Royal Statistical Society, B* **34**, pp. 187-200.

Cox, D.R. (1962): *Renewal Theory*, Methuen and Co. Ltd. London.

Daley, D.J. and D. Vere-Jones (2002): *An Introduction to the Theory of Point Processes: Elementary Theory and Methods* Vol 1, Springer-Verlag, New York.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, **39**(1), pp. 1-38.

Droost, F.C. and B.J.M. Werker (2004): "Semiparametric duration models," *Journal of Business and Economic Statistics*, **22**, pp. 40-50.

Fernandes, M. and J. Grammig, (2006): "A family of autoregressive conditional duration models," *Journal of Econometrics*, **127**, pp. 35-68.

Feng, D., G.J. Jiang and P.X.-K. Song, (2004): "Stochastic conditional duration models with 'leverage' effect for financial transaction data," *Journal of Financial Econometrics*, **2**, pp. 390-421.

Focardi, S.M, and F.J. Fabozzi, (2005): "An autoregressive conditional duration model of credit-risk contagion," *The Journal of Risk Finance* **6**, No.3, pp. 208-225.

Engle, R.F. (2002): "New frontiers for ARCH models," *Journal of Applied Econometrics*, **17**, pp 425-446.

Engle, R.F. and J.R. Russell, (1998): "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, **66**, No. 5, pp. 1127-1162.

Gagliardini, P. and C. Gourieroux, (2005): " Stochastic migration models with application to corporate risk," *Journal of Financial Econometrics*, **3**(2), pp. 188-226.

Gagliardini, P. and C. Gourieroux, (2005): "Migration correlation: Definition and efficient estimation," *Journal of Banking & Finance*, **29**, pp. 865-894.

Gerhard, F. and N. Hautsch, (2001): "Semiparametric autoregressive conditional proportional hazard models," *Working Paper - Nuffield College*, Oxford University.

Gill, R.D. (2001): "Product integration," *Working paper - University of Utrecht.*

Godambe, V.P. (Ed. 1985): *Estimating Functions.* Oxford University Press, New York.

Goodman, G.S. and S. Johansen (1973): "Kolmogorov's differential equations for non-stationary, countable state Markov processes with uniformly continuous transition probabilities," *Proceedings of the Cambridge Philosophical Society* **73**, pp. 119-138.

Gouriéroux, C., A. Monfort and A. Trognon (1984): "Pseudo maximum likelihood methods: theory," *Econometrica*, **52**, pp. 681-700.

Gray, S. (1996): "Modelling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, **42**(1), pp. 27-62.

Grimmett, G. R. and D. R. Stirzaker, (1992): *Probability and Random Processes, 2^{nd} Edition.* Oxford University Press, New York.

Ghysels, E., C. Gourieroux and J. Jasiak (2004): "Stochastic volatility duration models," *Journal of Econometrics*, **119**, pp. 413-433.

Harvey, A.C., E. Ruiz and N. Shephard (1994): "Multivariate stochastic variance model ," *Review of Economic Studies*, **61**, pp. 247-264.

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

Hujer, R., S. Vuletic and S. Kokot (2002): "The Markov switching ACD model," *Working Paper Series: Finance & Accounting*, Frankfurt University.

Isaacson, D. L. and R. W. Madsen(1976): *Markov Chains Theory and Applications*, Wiley, New York.

Jarrow, R. A., D. Lando, and S. M. Turnbull, (1997): "A Markov model for the term structure of credit spreads," *Review of Financial Studies* **10**, Nr. 2 pp.

Jasiak, J. (1998): "Persistence in intertrade durations," *Finance*, **19**, pp. 166-195.

Kalbfleisch, J.D. and R.L. Prentice . (2002): *The Statistical Analysis of Failure Time Data, 2nd Ed.*, Wiley, New York.

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

Khinchin, A. Y. (1956): "On Poisson sequences of chance events," *Theory of Probability and its Applications*, **1**(3) pp. 291 -297.

Koopman, S.J., Lucas, A. and A. Monteiro (2008): "The multi-state latent factor intensity model for credit rating transitions," *Journal of Econometrics*,**142**(1) pp. 405-430.

Koulikov, D. (2002): "Modelling sequences of long-memory positive weakly stationary random variables," *William Davidson Institute Working Paper Nr. 493.*

Lévy, P. (1954): "Processus semi-Markoviens," *Proceedings from the International Congress of Mathematics*, **3**, pp. 416-426.

Lewis, P.A.W (Ed. 1972): *Stochastic Point Processes; Statistical Analysis, Theory and Application*, Wiley, New York.

Meitz, M. & T. Terasvirta, (2006): "Evaluating models of autoregressive conditional duration," *Journal of Business and Economic Statistics*, **24**, pp. 104-124.

Monteiro, A., G.V. Smirnov, and A. Lucas, 2006," Nonparametric estimation for non-homogeneous semi-Markov processes: An application to credit risk." *Tinbergen Institute Discussion Paper* TI06-024/2.

Ogata, Y. and K. Katsura (1986): "Point Process models with linearly parameterized intensity for the application to earthquake catalogue," *Journal of Applied Probability*, **23**(A), pp. 1735-1748.

Peiris, S., D. Allen and K.H. Ng (2008): "Estimation of ACD models using estimating functions: A simulation study," *Working Paper - University of Sydney.*

Pyke, R. (1961): "Markov renewal processes with finitely many states," *The Annals of Mathematical Statistics*, **32**, Nr 4, pp. 1243-1259.

Richard, J-F. and W. Zhang. (2007): "Efficient high-dimensional importance sampling," *Journal of Econometrics*, **141**, pp. 1385-1411.

Ruiz, E. (1994): "Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, **63** pp. 289-306.

Russell, J.R. (1999): "Econometric modeling of multivariate irregularly spaced high-frequency data," *Working Paper - University of Chicago.*

Smith, W.L. (1955): "Regenerative stochastic processes," *Proceedings of the Royal Society*, Ser. A, **232**, pp. 6-31.

Snyder, D. L. (1975): *Random Point Processes*, Wiley, New York.

Snyder, D. L., and Miller (1991): *Random Point Processes in Time and Space*, Springer-Verlag, New York.

Zhang, M.Y., R. Russell and R.S. Tsay (2001): "A nonlinear autoregressive conditional duration model with applications to financial transaction data," *Journal of Econometrics*, **104**, pp. 179-207.