Effective measure of endogeneity for the Autoregressive Conditional Duration point processes via mapping to the self-excited Hawkes process

V. Filimonov,* S. Wheatley,† and D. Sornette‡

Department of Management, Technology and Economics,

ETH Zürich, Scheuchzerstrasse 7, CH-8092 Zürich, Switzerland

(Dated: June 11, 2013)

In order to disentangle the internal dynamics from exogenous factors within the Autoregressive Conditional Duration (ACD) model, we present an effective measure of endogeneity. Inspired from the Hawkes model, this measure is defined as the average fraction of events that are triggered due to internal feedback mechanisms within the total population. We provide a direct comparison of the Hawkes and ACD models based on numerical simulations and show that our effective measure of endogeneity for the ACD can be mapped onto the "branching ratio" of the Hawkes model.

I. INTRODUCTION

An outstanding challenge in socio-economic systems is to disentangle the internal dynamics from the exogenous influence. It is obvious that any non-trivial system is both subject to external shocks as well as to internal organization forces and feedback loops. In a given observation set, it seems in general hopeless to separate the contributions resulting from external perturbations from internal fluctuations and responses. However, one would like to understand the interplay between endogeneity and exogeneity (the 'endo-exo' problem, for short) in order to characterize the reaction of a given system to external influences, to quantify its resilience, and explain its dynamics. Using the class of self-exciting conditional

*Electronic address: vfilimonov@ethz.ch

†Electronic address: wspencer@ethz.ch

‡Electronic address: dsornette@ethz.ch

Hawkes Poisson processes [1, 2], some progress has recently been made in this direction [3–6].

Indeed, in the modeling of complex point processes in natural and socio-economic systems, the Hawkes process [1, 2] has become the gold standard due to its simple construction and flexibility. Nowadays it is being successfully used for modeling sequences of triggered earthquakes [7]; genomic events along DNA [8]; spread of violence [9] and crime [10] across some regions; high frequency fluctuations of financial prices [11] and probabilities of credit defaults [12]. Being closely related to branching processes [13], the Hawkes model combines, in a natural and parsimonious way, exogenous influences with self-excited dynamics. It accounts simultaneously for the co-existence and interplay between the exogenous impact on the system and the endogenous mechanism where past events contribute to the probability of occurrence of events in future. Moreover, using the mapping of the Hawkes process onto a branching process, it is possible to construct a representation of the sequence of events according to a branching structure, with each event leading to a whole tree of offsprings.

The linear construction of the Hawkes model allows one to separate exogenous events and develop a single parameter, the so-called "branching ratio" η that directly measures the level of endogeneity in the system. The branching ratio can be interpreted as the fraction of endogenous events within the whole population of events [14, 15]. The branching ratio provides a simple and illuminating characterization of the system, in particular with respect to its fragility and susceptibility to shocks. For $\eta < 1$, on average, the proportion $1 - \eta$ of events arrive to the system externally, while the proportion η of events could be traced back to the influence of past dynamics. As η approaches 1 from below, the system becomes "critical", in the sense that its activity is mostly endogenous or self-fulfilling. More precisely, its activity becomes hyperbolically sensitive to external influences. The regime $\eta > 1$ corresponds to the occurrence of an unbounded explosion of activity nucleated by just a few external events (e.g., news) with non-zero probability. In any realistic case, when present, this explosion will be observable in finite time. Not only does the Hawkes model provide this valuable parameter η, but it also amenable to an easy and transparent estimation by maximum likelihood [16, 17] without requiring stochastic declustering [18, 19], which is essential in the branching processes' framework but has several limitations [20].

However, the Hawkes model is not the only model that describe self-excitation in point processes. In particular, the Autoregressive Conditional Duration (ACD) model [21, 22] and

the Autoregressive Conditional Intensity (ACI) model [23] has been introduced for econometric applications. A similar concept was used in the so-called Autoregressive Conditional Hazard (ACH) model [24]. These processes were designed to mimic properties of the famous models of Autoregressive Conditional Heteroskedasticity (ARCH) [25] and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) [26] that are being successfully used in econometrics to explain volatility clustering and self-excitation in price time series. Some other modifications of ACD models such as Fractionally Integrated ACD (FIACD) [27] or Augumented ACD (AACD) [28] were introduced to account for additional effects (such as long memory) or to increase the flexibility of the model (for a more detailed review, see [29] and references therein).

In general, all approaches to modeling self-excited point processes can be separated into the classes of Duration-based (represented by the ACD model and its derivations) and Intensity-based approaches (Hawkes, ACH, ACI, and so on), which define a stochastic expression for inter-event durations and intensity respectively. Of all the models, as discussed above, the Hawkes process dominates by far in the class of intensity-based model, and the ACD model – a direct offspring of the GARCH-family – is the most used duration-based model.

Despite belonging to different classes, both models describe the same phenomena and exhibit similar mathematical properties. In this article, we aim to establish a link between the ACD and Hawkes models. We show that, despite the fact that the ACD model cannot be directly mapped onto a branching process, and thus the branching ratio for this model cannot be derived, it is possible to introduce a parameter $\zeta \in [0, 1]$ that serves as an effective degree of endogeneity in the ACD model. We show that this parameter shares important properties with the branching ratio $\eta \in [0, 1]$ in the framework of the Hawkes model. Namely, both ζ and η characterize stationarity properties of the models, and provide an effective transformation of the exogenous excitation of the system onto its total activity. By numerical simulations, we show that there exists a monotonous relationship between the parameter ζ of the ACD model and the branching ratio η of the corresponding Hawkes model. In particular, the purely exogenous case ($\eta = 0$) and the critical state ($\eta = 1$) are exactly mapped to the corresponding values $\zeta = 0$ and $\zeta = 1$. We validate our results by goodness-of-fit tests and show that our findings are robust with respect to the specification of the memory kernel of the Hawkes model.

The article is structured as follows. In section II, we introduce the Hawkes and ACD models and briefly discuss their properties. Section III introduces the branching ratio and relates it to the measure of endogeneity within the framework of the Hawkes model. In section IV, we discuss similarities between the Hawkes and ACD models, and identify a parameter in the ACD model that can be treated as an effective degree of endogeneity. We support our thesis with extensive numerical simulations and goodness-of-fit tests. In section V, we conclude.

II. MODELS OF SELF-EXCITED POINT PROCESSES

Let us define a *univariate point process* of event times $\{t_i\}_{i\in\mathbb{N}_{>0}}$ ($t_i > t_j$ for $i > j$) with the *counting process* $\{N(t)\}_{t\geq 0} = \max(i : t_i \leq t)$, and the *duration process* of inter-event times $\{\delta t_i\}_{i\in\mathbb{N}_{>0}} = t_i - t_{i-1}$. Properties of the point process $\{t_i\}$ are usually described with the *(unconditional) intensity process* $\lambda(t) = \lim_{h\downarrow 0} \frac{1}{h}\Pr[N(t+h) - N(t) > 0]$ and *conditional intensity process* $\lambda(t|\mathcal{F}_{t-}) = \lim_{h\downarrow 0} \frac{1}{h}\Pr[N(t+h) - N(t) > 0|\mathcal{F}_{t-}]$, which is adapted to the natural *filtration* $\mathcal{F}_{t-} = (t_1, \ldots, t_i; \forall i < N(t))$ representing the history of the process.

The well-known *Poisson point process* is defined as the point process whose conditional intensity does not depend on the history of the process and is equal to a deterministic constant:

$$\lambda(t|\mathcal{F}_{t-}) \equiv \lambda(t) = \lambda_0 > 0, \tag{1}$$

The *non-homogenous Poisson process* extends expression (1) to account for time-dependence of both conditional and unconditional intensity functions: $\lambda(t|\mathcal{F}_{t-}) \equiv \lambda(t) = \lambda_0(t) > 0$. Both homogeneous and non-homogeneous Poisson processes are completely memoryless, which means that the durations $\{\delta t_i\}$ are independent from each other and are completely determined by the exogenous parameter (function) $\lambda_0(t)$.

The *Self-excited Hawkes process* and *Autoregressive Conditional Durations (ACD)* model, which are described in this article, extend the concept of the Poisson point processes by adding path dependence and non-trivial correlation structures. These models represent two different approaches in modelling point processes with memory: the so called *intensity-based* and *duration-based* approaches. As follows from their names, the first approach focuses on models for the conditional intensity function $\lambda(t|\mathcal{F}_{t-})$ and the second considers models of the durations $\{\delta t_i\}$. For example, in the context of the intensity-based approach, the

Poisson process is defined by equation (1). In the context of the duration-based approach, the Poisson process is defined as the point process whose durations $\{\delta t_i\}$ are independent and identically distributed (iid) random variables with exponential probability distribution function $f(\delta t) = \lambda_0 \exp(-\lambda_0 \delta t)$.

A. Hawkes Model

The linear Hawkes process [1, 2], which belongs to the class of intensity-based models, has its conditional intensity $\lambda(t|\mathcal{F}_{t-})$ being a stochastic process of the following general form:

$$\lambda(t|\mathcal{F}_{t-}) = \mu(t) + \int_{-\infty}^{t} h(t-s)dN(s), \tag{2}$$

where $\mu(t)$ is the *background intensity*, which is a deterministic function of time that accounts for the intensity of arrival of *exogenous* events (not dependent on history). A deterministic *kernel function* $h(t)$, which should satisfy causality ($h(t) = 0$ for $t < 0$), models the *endogenous* feedback mechanism (memory of the process). Given that each event arrives instantaneously, the differential of the counting process $dN(t)$ can be represented in the form of a sum of delta-functions $dN(t) = \sum_{t_i<t} \delta(t - t_i)dt$, allowing (2) to be rewritten in the following form:

$$\lambda(t|\mathcal{F}_{t-}) = \mu(t) + \sum_{t_i<t} h(t - t_i). \tag{3}$$

It can be shown (and we will discuss this point in the following section) that the stationarity of the process (3) requires that

$$0 < \int_0^\infty h(t)dt < 1. \tag{4}$$

The shape of the kernel function $h(t)$ defines the correlation properties of the process. In particular, the geophysical applications of the Hawkes model, or more precisely of its spatio-temporal extension called the *Epidemic-Type Aftershock sequence (ETAS)* [7, 30, 31], assume in general a power law time-dependence of the kernel $h(t)$:

$$h(t) = \frac{K}{(t+c)^\varphi}\chi(t), \tag{5}$$

that describes the modified Omori-Utsu law of aftershock rates [32, 33]. Financial applications [11, 15, 34, 35] traditionally use an exponential kernel

$$h(t) = a \exp(-t/\tau)\chi(t), \tag{6}$$

which has been originally suggested by Hawkes [1] and ensures Markovian properties of the model [36]. In both cases, a Heaviside function $\chi(t)$ ensures the validity of the causality principle. The stationarity condition (4) requires $Kc^{1-\varphi}/(\varphi-1) < 1$ for the power law kernel and $a\tau < 1$ for the exponential kernel.

In the present work, we focus on the Hawkes model with an exponential kernel (6) and background intensity $\mu(t)$ that does not depend on time: $\mu(t) \equiv \mu > 0$. We introduce a new dimensionless parameter, $\eta = a\tau$, which will be discussed in detail later, which allows us to write the final expression for the conditional intensity as follows:

$$\lambda(t|\mathcal{F}_{t-}) = \mu + \frac{\eta}{\tau}\sum_{t_i<t}\exp\left(-\frac{t-t_i}{\tau}\right) \ . \tag{7}$$

Then, the stationarity condition reads $\eta < 1$. In order to check the robustness of the results presented below, in particular with respect to the choice of the memory kernel, we have also considered a power law kernel (5) with time-independent background intensity $\mu(t) \equiv \mu > 0$. Similarly to the exponential kernel, the integral from 0 to $+\infty$ of the memory kernel defines the dimensionless parameter $\eta = Kc^{1-\varphi}/(\varphi-1)$, which allows us to rewrite the Hawkes model with power law kernel as:

$$\lambda(t|\mathcal{F}_{t-}) = \mu + \eta c^{1-\varphi}(\varphi-1)\sum_{t_i<t}\frac{1}{(t-t_i+c)^{\varphi}}, \tag{8}$$

Again, the stationarity condition reads $\eta < 1$.

B. Autoregressive Conditional Durations (ACD) Model

The class of *Autoregressive Conditional Durations* (ACD) models has been introduced by Engle and Russell [21, 22] in the field of econometrics to model financial data at the transaction level. The ACD model applies the ideas of the Autoregressive Conditional Heteroskedasticity (ARCH) [25] model, which separates the dynamics of a stationary random process into a multiplicative random error term and a dynamical variance that regresses the past values of the process. In the spirit of the ARCH, the ACD model is represented by the duration process δt_i in the form

$$\delta t_i = \psi_i \epsilon_i, \tag{9}$$

where ϵ_i defines an iid random non-negative variable with unit mean $\mathrm{E}[\epsilon_i] = 1$, and the function $\psi_i \equiv \psi(N(\mathcal{F}_{t-});\theta)$ is the conditional expected duration: $\mathrm{E}[\delta t_i|\mathcal{F}_{t-}] = \psi_i$. Here, θ

represents the set of parameters of the model. From expression (9), one can simply derive the conditional intensity of the process [22]:

$$\lambda(t|\mathcal{F}_{t-}) = \lambda_\epsilon \left(\frac{t - t_{N(t)}}{\psi_{N(t)+1}} \right) \frac{1}{\psi_{N(t)+1}}, \tag{10}$$

where $\lambda_\epsilon(s)$ represents the intensity function of the noise term, ϵ_i. Assuming ϵ_i to be iid exponentially distributed, one can call this model (9) an *Exponential ACD* model.

The conditional expected duration $\psi(N(\mathcal{F}_{t-}); \theta)$ of the ACD(p,q) model, where (p, q) denotes the order of the model, is defined as an autoregressive function of the past observed durations δt_i and the conditional durations ψ_i themselves:

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \delta t_{i-j} + \sum_{k=1}^{q} \beta_k \psi_{i-j}, \tag{11}$$

where $\omega > 0$, $\alpha_j \geq 0$ and $\beta_k \geq 0$ are parameters of the model that constitute the set $\theta = \{\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_p\}$. The stationarity condition for the ACD model has the form [22]:

$$\sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k < 1. \tag{12}$$

In the simple ACD(1,1) case that is considered in the present article, equation (11) is reduced to:

$$\psi_i = \omega + \alpha \delta t_{i-1} + \beta \psi_{i-1}. \tag{13}$$

Similarly, the conditional intensity (10) of the Exponential ACD(1,1) has the form:

$$\lambda(t|\mathcal{F}_{t-}) = \frac{1}{\psi_{N(t)+1}} = \frac{1}{\omega + \alpha \delta t_{N(t)} + \beta \psi_{N(t)}} \tag{14}$$

and the stationarity condition (12) reduces to $\alpha + \beta < 1$.

III. THE BRANCHING RATIO AS A MEASURE OF ENDOGENEITY IN THE HAWKES MODEL

The linear structure of the Hawkes process (3) allows one to consider it as a cluster process in which the random process of cluster centers $\{t_i^{(c)}\}_{i \in \mathbb{N}_{>0}}$ is the Poisson process with rate $\mu(t)$. All clusters associated with centers $\{t_i^{(c)}\}$ are mutually independent by construction and can be considered as a *generalized branching process* [37], illustrated in fig. 1.

In this context, each event $\{t_i\}$ can be either an *immigrant* or a *descendant*. The rate of immigration is determined by the background intensity $\mu(t)$ and results in an exogenous random process. Once an immigrant event occurs, it generates a whole cluster of events. Namely, a zeroth-order event (which we will call the *mother event*) can trigger one or more first-order events (*daughter events*). Each of these daughters, in turn, may trigger several second-order events (the grand-daughters of the initial mother), and so on. All first-, second- and higher-order events form a cluster and are called descendants (or *aftershocks*) and represent endogenously driven events that appear due to internal feedback mechanisms in the system. It should be noted that this mapping of the Hawkes process (3) onto the branching process (fig. 1) is possible only due to the linearity of the model, and is not valid for nonlinear self-excited point processes, such as the class of nonlinear mutually excited point processes [38], of which the Multifractal stress activation model [39] and the self-excited multifractal model [40] are particular implementations.

The crucial parameter of the branching process is the *branching ratio* (n), which is defined as the average number of daughter events per mother event. Depending on the branching ratio, there are three regimes: (i) *sub-critical* ($n < 1$), (ii) *critical* ($n = 1$) and (iii) *super-critical* or explosive ($n > 1$). Starting from a single mother event (or immigrant) at time t_1, the process dies out with probability 1 in the sub-critical and critical regimes and has a finite probability to explode to an infinite number of events in the super-critical regime. The critical regime for $n = 1$ separates the two main regimes and is characterized by power law statistics of the number of events and in the number of generations before extinction [41]. For $n \leq 1$, the process is stationary in the presence of a Poissonian or more generally stationary flux of immigrants.

Being the parameter that describes the clustering structure of the branching process, the branching ratio n defines the relative proportion of exogenous events (immigrants) and endogenous events (descendants or aftershocks). Moreover, in the sub-critical regime, in the case of a constant background intensity ($\mu(t) = \mu = \text{const}$), the branching ratio is exactly equal to the fraction of the average number of descendants in the whole population [14]. In other words, the branching ratio is equal to the proportion of the average number of endogenously generated events among all events and can be considered as an effective measure of endogeneity of the system.

To see this, let us count separately the rates of exogenous and endogenous events. The

rate of exogenous immigrants (zeroth-order events) is equal to the background activity rate: $R_{exo} = \mu$. Each immigrant independently gives birth, on average, to n daughters and thus the rate of first-order events is equal to $r_1 = \mu n$. In turn, each first-order event produces, on average, n second-order events, whose rate is equal to $r_2 = nr_1 = \mu n^2$. Continuing this process ad infinitum and summing over all generations, we obtain the rate of all endogenous descendants:

$$R_{endo} = \sum_{i=1}^{\infty} r_i = \mu \sum_{i=1}^{\infty} n^i = \frac{\mu n}{1-n}, \tag{15}$$

which is finite for $n < 1$. The global rate is the sum of the rates of immigrants and descendants and equal to

$$R = R_{exo} + R_{endo} = \mu + \frac{\mu n}{1-n} = \frac{\mu}{1-n}. \tag{16}$$

And the proportion of descendants (endogenously driven events) in the whole system is equal to the branching ratio:

$$\frac{R_{endo}}{R} = n. \tag{17}$$

Calibrating n on the data therefore provides a direct quantitative estimate of the degree of endogeneity.

In the framework of the Hawkes model (3) with $\mu(t) = \mu = \text{const}$, the branching ratio n is easily defined via the kernel $h(t)$:

$$n = \int_0^{\infty} h(t)dt. \tag{18}$$

For the exponential parametrization (6), the branching ratio, $n = a\tau$, is equal to a dimensionless parameter $n \equiv \eta$ previously introduced. The Hawkes framework provides a convenient way of estimating the branching ratio, $n \equiv \eta$, from the observations $\{t_i\}$, using the Maximum Likelihood method, which benefits from the fact that the log-likelihood function is known in closed form for Hawkes processes [16, 17].

IV. THE EFFECTIVE DEGREE OF ENDOGENEITY IN THE AUTOREGRESSIVE CONDITIONAL DURATIONS (ACD) MODEL

A. Formal similarities between the ACD and Hawkes models

Note that the ACD(p,q) and Hawkes models operate on different variables with inverse dimensions: duration δt for the ACD(p,q) model and conditional intensity $\lambda(t|\mathcal{F}_{t-})$ for the

Hawkes model, which is of the order of the inverse $1/\delta t$ of the duration δt. As a consequence, equations (20) and (16) apply to different statistics (average durations $\mathrm{E}[\delta t]$ and average rate $R = \mathrm{E}[1/\delta t]$). Moreover, the ACD model cannot be exactly mapped onto a branching process.

Nevertheless, the ACD model shares many similarities with the Hawkes model and their point processes exhibit similar degrees of clustering. In particular, for the ACD defined by expression (11), the combined parameter,

$$\zeta = \sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k \tag{19}$$

plays a similar role to the parameter η of the Hawkes process with an exponential kernel (7). The similarities start with the stationarity conditions (4) and (12) which require $\eta < 1$ for the Hawkes model and $\zeta < 1$ for the ACD, but go much deeper than the simple idea of "effective distance" to a non-stationary regime.

As we have seen in the previous section, η defines the effective degree of endogeneity (17) that translates the exogenous rate $R_{exo} = \mu$ into the total rate $R_{total} = R_{exo}/(1-\eta)$. Similarly, let us study the role of endogenous feedback in the ACD model. For $\alpha_j = \beta_k = \zeta = 0$, the ACD(p,q) model (9),(11) reduces to a simple Poisson process with durations $\delta t_i = \omega \epsilon_i$ having an average value of $\mathrm{E}[\delta t_i] = \omega$, which can be considered as the exogenous factor. When $\alpha_j > 0$ and $\beta_k > 0$, there is an amplification of the average durations. Considering the average of eq. (11) in the stationary regime ($\mathrm{E}[\delta t_{i-1}] = \mathrm{E}[\delta t_i]$ and $\mathrm{E}[\psi_{i-1}] = \mathrm{E}[\psi_i]$), and taking into account eq. (9), we obtain the following expression for the mean duration in the stationary regime:

$$\mathrm{E}[\delta t] = \frac{\omega}{1 - \sum_{j=1}^{p} \alpha_j - \sum_{k=1}^{q} \beta_k} \equiv \frac{\omega}{1-\zeta}. \tag{20}$$

Equations (20) and (16) share the same functional dependence, with a divergence when the corresponding control parameters η and ζ approach 1.

B. Empirical dependence of the effective branching ratio $\hat{\eta}$ as a function of $\zeta = \alpha + \beta$ for the ACD(1,1) process

In order to quantify the similarities between the ACD and Hawkes models outlined in the previous section, we have performed the following numerical study. We simulated realizations

of the ACD(1,1) process and calibrated the Hawkes model to it. The traditional way of fitting the Hawkes model uses the maximum likelihood method [17], which is asymptotically normal and asymptotically efficient [16]. We have used the R package "PtProcess" [42], which provides a convenient framework for Hawkes models (3) with arbitrary kernel $h(t)$ and background intensity $\mu(t)$. Then, we maximized the likelihood function using a Newton-type non-linear maximization [43, 44]. The Appendix reports a study of the finite sample bias and efficiency of the Hawkes maximum likelihood estimator. We find that the estimation error $|\hat{\eta} - \eta|$ of the branching ratio (without model error) measured with the 90% quantile ranges does not exceed 0.1 for all values $\eta \leq 0.9$.

More precisely, we want to quantify similarities between control parameters ζ of the model ACD(p,q) and η of the exponential Hawkes model. For this, we have simulated realizations of the ACD process and estimated the parameter η from these realizations. The parameter ω of the ACD(p,q) (11) model defines the time scale. Without loss of generality, we let $\omega = 1$, which accounts for a linear transformation of time $\tilde{t}_i = t_i/\omega$ in equations (9) and (11). For the sake of simplicity, we present our results for the ACD(1,1) model, for which the dimensionless parameter ζ reduces to $\zeta = \alpha + \beta$. However, our findings are robust to the choice of the order of the ACD model and can be easily generalized to the case of $p, q > 1$. The parameters α and β were chosen so that $\zeta = \alpha + \beta$ spanned $[0, 1]$ at 40 equidistant points. For each of the 40 values of ζ, we have generated 100 realizations of the corresponding exponential ACD(1,1) process. Each realization of 3500 events was generated by a recursive algorithm using eq. (13). In order to minimize the impact of edge effects, the first 500 points of each realization were discarded. Then, the Hawkes model (7) was calibrated on these synthetic datasets.

For each calibration, we have performed a goodness-of-fit test based on residual analysis [7], which consists of studying the so-called residual process defined as the nonparametric transformation of the initial time-series t_i into

$$\xi_i = \int_0^{t_i} \hat{\lambda}(t|\mathcal{F}_{t-})dt, \tag{21}$$

where $\hat{\lambda}(t|\mathcal{F}_{t-})$ is the conditional intensity of the Hawkes process (7) estimated with the maximum likelihood method. Under the null hypothesis that the data has been generated by the Hawkes process (7), the residual process ξ_i should be Poisson with unit intensity [45]. Visual analysis involves studying the cusum plot or Q-Q plot and may be complemented

with rigorous statistical tests. Under the null hypothesis (Poisson statistics of the residual process ξ_i), the inter-event times in the residual process, $\delta\xi_i = \xi_i - \xi_{i-1}$, should be exponentially distributed with CDF $F(\delta\xi) = 1 - \exp(-\delta\xi)$. Thus, the random variables $U_i \equiv F(\delta\xi_i) = 1 - \exp(-\delta\xi_i)$ should be uniformly distributed in $[0, 1]$. We have performed rigorous Kolmogorov-Smirnov tests for uniformity and provided the corresponding p-values.

We start with a visual comparison of realizations generated with the two models. Fig. 2 presents a comparison of the conditional intensities and durations for (i) simulations of the ACD(1,1) process and, (ii) simulations of the Hawkes process with parameters calibrated to the corresponding ACD process realization. Visual similarities are striking for all four ACD-Hawkes pairs: Total, average, and maximum durations are similar. Moreover, bursts of short and long durations are of similar length. The conditional intensities fluctuate in a similar range and show qualitatively similar clustering of events, although the ACD conditional intensity is constant between events while the Hawkes decays exponentially. Quantitatively, the distributions of durations also show a large degree of similarity.

Fig. 2 also reveals one important property of the ACD model. Despite the fact that many statistical properties (such as average durations (20)) are defined by the control parameter $\zeta = \alpha + \beta$, α and β have different impacts on the effective degree of endogeneity η. For instance, case (B) $\alpha = 0.38, \beta = 0.13$, and case(C) $\alpha = 0.13, \beta = 0.38$ both have the same $\zeta = 0.51$ but $\hat{\eta} = 0.52$ for (B) and $\hat{\eta} = 0.22$ for (C). The smaller endogeneity found in case (C) is compensated by a higher rate of exogenous events ($\hat{\mu} = 0.38$ for (C) compared with $\hat{\mu} = 0.24$ for (B)), resulting in a "flatter" conditional intensity for (C).

In order to explore this effect in simulations of the ACD model, for each value of $\zeta = \alpha+\beta$, we considered different relations between α and β: (i) $\alpha = \beta$ $(= \zeta/2)$, (ii) $\beta = 0$ $(\alpha = \zeta)$, (iii) $\alpha = 0$ $(\beta = \zeta)$, (iv) $\alpha = 3\beta$ $(= 3\zeta/4)$ and (v) $\beta = 3\alpha$ $(= 3\zeta/4)$. Fig. 3 presents the results of the fitting of the Hawkes model on realizations of the ACD(1,1) model in these five cases. The first striking observation is the existence of two fundamentally different behaviors observed for $\alpha = 0$ (case (iii)) versus $\alpha > 0$ (cases (i),(ii),(iv),(v)). For $\alpha = 0$, the estimated effective branching ratio $\hat{\eta}$ is 0 for all values of the control parameter $\zeta = \beta$, as shown in Figure 3B. This diagnoses a completely exogenous dynamics of the ACD process, which is indeed the expected diagnostic given that, for $\alpha = 0$, eq. (9) and (13) reduce to

$$\delta t_i = \psi_i \epsilon_i, \quad \psi_i = \omega + \beta \psi_{i-1}, \tag{22}$$

for which the dynamics of the conditional durations $\{\psi_i\}$ is purely deterministic and independent of the realized durations δt_i, while the later are entirely driven by the random term ϵ_i.

For $\alpha > 0$, we find similar non-trivial results. Fig. 3(B) shows the effective branching ratio $\hat{\eta}$ as a monotonously increasing function of ζ for all combinations of $\alpha \neq 0$ and β. In cases (ii) ($\beta = 0$) and (iv) ($\alpha = 3\beta$), the dependence of $\hat{\eta}$ on ζ is almost linear for $\zeta < 0.5$ and $\zeta < 0.9$ respectively and, for higher values of ζ, the convexity increases. In case (i) ($\alpha = \beta$), $\hat{\eta}$ depends linearly on ζ for $\zeta > 0.4$ with a very good approximation. Finally, in case (v) ($\alpha = 3\beta$), the curvature of $\hat{\eta}(\zeta)$ is significant over the range of $0.3 < \zeta < 0.9$. Remarkably, all four dependencies converge to the same value $\hat{\eta} \approx 0.9$ for $\zeta = 1$.

Figure 4 presents the dependence of the effective branching ratio $\hat{\eta}$ on the control parameter $\zeta = \alpha + \beta$ after correction of the bias in estimation due to finite size effects presented in the Appendix and summarized in fig. 7). All dependencies of $\hat{\eta}$ as a function of ζ converge to the critical value $\zeta = 1$.

Figure 5 generalizes fig. 3(B) by presenting the dependence of the effective branching ratio $\hat{\eta}$ (corrected for the finite sample bias determined in the Appendix) on the parameters α and β separately. As expected, the impact of a change of α is much larger than that of β. There is a region, delineated by the dashed line, within which the Hawkes model is rejected at the 5% level for the Kolmogorov-Smirnov test. For most combinations of α and β such that $0.6 \lesssim \alpha + \beta \lesssim 0.95$, the Hawkes model is rejected. Interestingly, the Hawkes model is not rejected in the case where β is kept significantly larger than α, and it is only rejected in a small interval in the extreme opposite case where $\beta \equiv 0$. The model is often not rejected for large values of $\hat{\eta}$.

C. Differences between the ACD and Hawkes models

Despite similarities, the Hawkes and ACD models exhibit some important differences. Fig. 3A shows that the effective background rate $\hat{\mu}$ estimated by the Hawkes model is a decreasing function of the control parameter ζ. This is an indirect consequence of the dependence of the expected duration on ζ given by expression (20). In contrast to the Hawkes model (2), for which the background rate $\mu(t)$ completely describes the exogenous impact on the system, the parameter ω of the ACD model (11) is not the only factor

embodying the exogenous activity and there is no strict decoupling between the exogenous driver ω and endogenous level ζ as occurs for the parameters μ and η of the Hawkes model. In other words, in contrast to the Hawkes model, the ACD in its classical form (11) does not provide a clean distinction between exogenous and endogenous activities.

Another difference between the Hawkes and ACD models can be observed in fig. 3(D), which presents a residual analysis of the calibration of realizations of the ACD process by the Hawkes model using the Kolmogorov-Smirnov test. The null hypothesis that the realizations of the ACD process are generated by the Hawkes model is rejected at the 5% confidence level for $\zeta > 0.6$ in case (i) ($\alpha = \beta$). For case (ii) ($\beta = 0$) and (iv) ($\alpha = 3\beta$), the null hypothesis is rejected for even lower $\zeta > 0.4$. However, for case (v) ($\beta = 3\alpha$), the null cannot be rejected for almost all values of the control parameter ζ, except for a small interval around $\zeta \approx 0.8$.

D. Influence of the memory kernel $h(t)$ of the calibrating Hawkes process

Finally, we need to discuss the choice of the kernel $h(t)$ in the specification of the Hawkes model (3) used in the calibration of the realizations generated with the ACG process. The use of the exponential kernel (6) is a priori justified by the short memory of the ACD(1,1) process. Indeed, the autocorrelation function of the ACD(1,1) model decays exponentially [29], and the same can be shown explicitly for the GARCH(1,1) model [46]. The choice of a short-memory exponential kernel for the Hawkes model ensures Markovian properties with a fast decaying autocorrelation function of the durations [36]. High p-values of the goodness-of-fit tests for parameters $\zeta < 0.5$ (see fig. 3) confirm the good mapping between the exponential ACD(1,1) and Hawkes processes with an exponential kernel.

In order to further validate the selection of the exponential kernel of the Hawkes process, we have compared the calibrations of realizations generated with the ACG process with the Hawkes model with the exponential kernel (6) and with the power law kernel (5). Since these models have a different number of parameters ($k = 3$ and $k = 4$ respectively), we compare them using the Akaike information criterion (AIC) [47]. The AIC is by far the most popular model comparison criterion used in the point process literature [48]. The AIC penalizes complex models by discounting the likelihood function L by the number k of parameters of the model. Specifically, the AIC suggests selecting the model with a minimum AIC value, where $AIC = 2k - 2\log L$.

TABLE I: Estimated parameters of the Hawkes model with exponential (6) and power law (5) kernels together with values of log-likelihood ($\log L_{exp}$ and $\log L_{pow}$) and Akaike information criterion (AIC_{exp} and AIC_{pow}) for cases presented in figure 2. Bold font identifies the lowest AIC value among the two models.

	α	β	$\theta_{H,exp} = (\hat{\mu}, \hat{n}, \hat{\tau})$	$\theta_{H,pow} = (\hat{\mu}, \hat{n}, \hat{c}, \hat{\varphi})$	$\log L_{exp}$	$\log L_{pow}$	AIC_{exp}	AIC_{pow}
A	0.05	0.05	(0.84, 0.07, 4.3)	(0.83, 0.10, 262.77, 73.13)	−343.0	−343.4	**692.0**	694.8
B	0.38	0.13	(0.24, 0.52, 5.6)	(0.21, 0.54, 501.14, 113.41)	−490.4	−491.4	**986.8**	990.8
C	0.13	0.38	(0.38, 0.22, 7.9)	(0.36, 0.23, 816.41, 105.17)	−509.3	−509.3	**1023.6**	1026.6
D	0.45	0.45	(0.02, 0.83, 28.4)	(0.01, 0.87, 203.91, 7.70)	−900.0	−901.5	**1806.0**	1811.0

Table I gives the results for the realizations presented in fig. 2. In terms of likelihood, the exponential and power law kernels give practically identical values ($\log L_{exp} \approx \log L_{pow}$). Penalizing model complexity with the AIC widens the gap, and the exponential kernel with one fewer parameters is selected under the AIC.

Notwithstanding their apparent strong difference, the estimated background intensities ($\hat{\mu}$) and branching ratios ($\hat{n}$) are almost the same for both memory kernels. This can explained by the fact that the parameters $\hat{\varphi}$ and $\hat{c}$ estimated for the power law kernel (5) are such that the later remains very close to an exponential kernel over a large time interval, as illustrated by figure 6 for case C ($\alpha = 0.13$, $\beta = 0.38$), which presents a direct comparison between the exponential kernel $\tilde{h}(t) = h(t)/\eta = \tau^{-1}\exp(-t/\tau))$ and the power law kernel $\tilde{h}(t) = c^{1-\varphi}(\varphi-1)(t+c)^{-\varphi}$. The corresponding ML estimates of their parameters are respectively $\hat{\tau} = 7.76$, $\hat{\varphi} = 105.17$ and $\hat{c} = 816.41$. The large value of the estimated exponent $\hat{\varphi}$ (of the order of 100) implies a fast decay, similar to an exponential function. Correlatively, the large value of the constant $\hat{c}$ implies the absence of the hyperbolic range (or "long tail") as well. The almost perfect coincidence is observed for up to times $t \simeq 30$, over which the kernels $\tilde{h}(t)$ decay by a factor of almost 50. For $t > 50$ for which the relative difference between the two kernels exceed 20%, the absolute values of $h(t)$ is less than $2 \cdot 10^{-4}$ so that the contribution of time scales beyond $t = 50$ to the total intensity (7),(8) becomes insignificant.

V. CONCLUSION

The present article positions itself within the financial and neoclassical econometrics literature that investigates the nature of the mechanisms that drive financial prices. The benchmark, called the "Efficient Market Hypothesis" (EMH) [49, 50], holds that markets only reacts to external inputs (information flow) and almost instantaneously reflect these inputs in the price dynamics. This purely exogenous view on price formation has been contradicted by many empirical observations (see for instance the original works [51, 52] and more recent [53, 54]), which show that only a minor percent of price movements can be explained by the relevant news releases. This implies a significant role for internal feedback mechanisms. Using the framework of Hawkes processes, two of us [15] have used the corresponding branching ratio to provide what is, to the best of our knowledge, the first quantitative estimate of the degree of endogeneity in financial markets. This degree of endogeneity is measured as the proportion of price moves resulting from endogenous interactions among the total number of all price moves (including both endogenous interactions and exogenous news).

In this context, the present article expands the quantification of endogeneity to the class of Autoregressive Conditional Duration (ACD) point processes. This is done by the introduction of the composite parameter ζ (19) associated with the parameters α_j and β_k, which control the dependence of the conditional expectated duration between events as a function of past realized duration and past conditional expected duration. We have shown that the parameter ζ can be mapped onto the branching ratio η that directly measures the level of endogeneity within the framework of the Hawkes self-excited conditional Poisson model. This result leads to a novel interpretation of the various econometric studies that analyzed high-frequency financial data with the ACD model.

An important conclusion derives from our mapping of the ACD onto the Hawkes process. Both original works [22, 23, 55] as well as more recent studies reviewed in Refs. [56, 57] have reported estimated parameters α_j and β_k that combine to extremely large values of ζ, often larger than 0.5, and up to 0.95. From the perspective offered by the present work and in particular from the mapping of ζ onto η, these empirical findings provide strong support to the hypothesis of a dominant endogenous or "reflexive" [58] component in the dynamics of financial markets.

The present work offers itself to a natural extension beyond point processes to the class of discrete time processes. There are several successful models of self-excitation within a discrete time framework, such as AR (auto-regressive), ARMA (auto-regressive moving average) [59] and GARCH models [26] and their siblings, as well as the recently introduced Self-Excited Multifractal (SEMF) model [40]. However, until now, there has been no framework that provides a direct quantification and estimation of the degree of endogeneity present in a given time series. As discussed above, the ACD(p,q) model in fact belongs to the class of GARCH(p,q) models, though not with normally distributed innovations but instead with iid distributed innovations with a Poisson distribution. By extension, this suggests a direct application of our present findings to GARCH models. This correspondence will benefit from the elaborate toolbox of calibration methods and the detailed accumulated knowledge of the statistical properties of GARCH models [46, 60].

Appendix: Finite sample bias of the Hawkes maximum likelihood estimator

In order to optimize the calibration of the Hawkes model on the ACD(1,1), we study the finite sample bias and efficiency of the Hawkes maximum likelihood estimator. For this, we have simulated realizations of the Hawkes process with a modified thinning procedure [61, 62] implemented in the same "PtProcess" package [42], and afterwards calibrated the Hawkes model on this synthetic data. It should be noted that simulation (and fitting [22]) of the ACD model is computationally easier than for the Hawkes model. Indeed, simulation of the Hawkes process with the thinning algorithm has complexity of $\mathrm{O}(N^2)$ (with possibility to reduce to $\mathrm{O}(N \log N)$ [63, 64]), compared with complexity of only $\mathrm{O}(N)$ for the ACD(1,1) model.

We swept the parameter η in the range $[0, 1]$, fixing other parameters to $\mu = 1$ and $\tau = 1$. We generated 100 realizations of the Hawkes process of size 3500 each. To reduce the edge effects of the thinning algorithm, we discarded the first 500 points of each realization and afterwards calibrated the parameters of the Hawkes model on these realizations of length 3000. Fig. 7 illustrates the bias and efficiency of the maximum likelihood estimator in our framework. The definition of the Hawkes model (3) requires the kernel $h(t)$ to be always positive. This implies $\eta \geq 0$, so the estimation of η is expected to have positive bias for small values, as seen in fig. 7. On the other hand, when η approaches the critical value of 1 from below, the memory of the system increases dramatically and, for critical state of $\eta = 1$, the memory becomes infinite. Thus, for a realization of limited length, the finite size will play a very important role and will result in a systematic negative bias for $\eta \lesssim 1$. This reasoning is supported by the evidence presented in figure 7, where one observes large systematic bias for $\eta > 0.9$. For values of the branching ratio not too close to 0 or 1, the bias is very small for almost all reasonable realization lengths (longer than 200 to 400 points). We also find that the bias for η close to 1 strongly depends on the realization length. Finally, fig. 7 illustrates the high efficiency of the maximum likelihood estimator: for values of $\eta < 0.9$, the estimation error $|\hat{\eta} - \eta|$ measured with the 90% quantile ranges does not exceed 0.1.

[1] A. G. Hawkes, Spectra of some self-exciting and mutually exciting point processes, Biometrika 58 (1) (1971) 83–90.

[2] A. G. Hawkes, Point Spectra of Some Mutually Exciting Point Processes, Journal of the Royal Statistical Society. Series B (Methodological) 33 (3) (1971) 438–443.

[3] D. Sornette, F. Deschâtres, T. Gilbert, Y. Ageon, Endogenous Versus Exogenous Shocks in Complex Networks: An Empirical Test Using Book Sale Rankings, Physical Review Letters 93 (22) (2004) 228701.

[4] F. Deschâtres, D. Sornette, Dynamics of book sales: Endogenous versus exogenous shocks in complex networks, Physical Review E 72 (1) (2005) 016112.

[5] D. Sornette, Endogenous versus Exogenous Origins of Crises, in: S. Albeverio, V. Jentsch, H. Kantz (Eds.), Extreme events in nature and society, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 95–119.

[6] R. Crane, D. Sornette, Robust dynamic classes revealed by measuring the response function of a social system, Proceedings of the National Academy of Sciences of the United States of America 105 (41) (2008) 15649–15653.

[7] Y. Ogata, Statistical models for earthquake occurrences and residual analysis for point processes, Journal of the American Statistical Association 83 (401) (1988) 9–27.

[8] P. Reynaud-Bouret, S. Schbath, Adaptive estimation for Hawkes processes; application to genome analysis, The Annals of Statistics 38 (5) (2010) 2781–2822.

[9] E. Lewis, G. O. Mohler, P. J. Brantingham, A. L. Bertozzi, Self-exciting point process models of civilian deaths in Iraq, Security Journal 25 (2012) 244–264.

[10] G. O. Mohler, M. B. Short, P. J. Brantingham, G. E. Tita, F. P. Schoenberg, Self-Exciting Point Process Modeling of Crime, Journal of the American Statistical Association 106 (493) (2011) 100–108.

[11] C. G. Bowsher, Modelling security market events in continuous time: Intensity based, multivariate point process models, Journal of Econometrics 141 (2) (2007) 876–912.

[12] E. Errais, K. Giesecke, L. R. Goldberg, Affine Point Processes and Portfolio Credit Risk, SIAM Journal on Financial Mathematics 1 (1) (2010) 642.

[13] T. E. Harris, The Theory of Branching Processes, Dover Phoenix Editions, 2002.

[14] A. Helmstetter, D. Sornette, Importance of direct and indirect triggered seismicity in the ETAS model of seismicity, Geophysical Research Letters 30 (11) (2003) 1576.

[15] V. Filimonov, D. Sornette, Quantifying reflexivity in financial markets: Toward a prediction of flash crashes, Physical Review E 85 (5) (2012) 056108.

[16] Y. Ogata, The asymptotic behaviour of maximum likelihood estimators for stationary point processes, Annals of the Institute of Statistical Mathematics 30 (1) (1978) 243–261.

[17] T. Ozaki, Maximum likelihood estimation of Hawkes' self-exciting point processes, Annals of the Institute of Statistical Mathematics 31 (1) (1979) 145–155.

[18] J. Zhuang, Y. Ogata, D. Vere-Jones, Stochastic declustering of space-time earthquake occurrences, Journal of the American Statistical Association 97 (458) (2002) 369–380.

[19] D. Marsan, O. Lengline, Extending Earthquakes' Reach Through Cascading, Science 319 (5866) (2008) 1076–1079.

[20] D. Sornette, S. Utkin, Limits of declustering methods for disentangling exogenous from endogenous events in time series with foreshocks, main shocks, and aftershocks, Physical Review E 79 (6) (2009) 061110.

[21] R. F. Engle, J. R. Russell, Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model, Journal of Empirical Finance 4 (2-3) (1997) 187–212.

[22] R. F. Engle, J. R. Russell, Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, Econometrica: Journal of the Econometric Society 66 (5) (1998) 1127–1162.

[23] J. R. Russell, Econometric Modeling of Multivariate Irregularly-Spaced High-Frequency Data, Working Paper, University of Chicago (1999) 1–40.

[24] J. D. Hamilton, O. Jorda, A Model for the Federal Funds Rate Target, Journal of Political Economy 110 (2002) 1135–1167.

[25] R. F. Engle, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica: Journal of the Econometric Society 50 (4) (1982) 987–1007.

[26] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (3) (1986) 307–327.

[27] J. Jasiak, Persistence in intratrade durations, Financial Analysts Journal 19 (1998) 166–195.

[28] M. Fernandes, J. Grammig, A family of autoregressive conditional duration models, Journal of Econometrics 130 (1) (2006) 1–23.

[29] L. Bauwens, N. Hautsch, Modelling Financial High Frequency Data Using Point Processes, in: T. Mikosch, J.-P. Kreiß, R. A. Davis, T. G. Andersen (Eds.), Handbook of Financial Time Series, Springer, 2009, pp. 953–979.

[30] D. Vere-Jones, T. Ozaki, Some examples of statistical estimation applied to earthquake data I. Cyclic Poisson and self-exciting models, Annals of the Institute of Statistical Mathematics 34 (1) (1982) 189–207.

[31] D. Vere-Jones, Stochastic Models for Earthquake Occurrence, Journal of the Royal Statistical Society. Series B (Methodological) 32 (1) (1970) 1–62.

[32] T. Utsu, A statistical study of the occurrence of aftershocks, Geophysical Magazine 30 (1961) 521–605.

[33] T. Utsu, Y. Ogata, The centenary of the Omori formula for a decay law of aftershock activity, Journal of Physics of the Earth 41 (1) (1995) 1–33.

[34] P. Hewlett, Clustering of order arrivals, price impact and trade path optimisation, In Workshop on Financial Modeling with Jump processes, Ecole Polytechnique.

[35] R. Cont, Statistical Modeling of High Frequency Financial Data: Facts, Models and Challenges, IEEE Signal Processing 28 (5) (2011) 16–25.

[36] D. Oakes, The Markovian Self-Exciting Process, Applied Probability Trust 12 (1) (1975) 69–77.

[37] A. G. Hawkes, D. Oakes, A Cluster Process Representation of a Self-Exciting Process, Journal of Applied Probability 11 (3) (1974) 493–503.

[38] P. Brémaud, L. Massoulié, Stability of nonlinear Hawkes processes, The Annals of Probability 24 (3) (1996) 1563–1588.

[39] D. Sornette, G. Ouillon, Multifractal Scaling of Thermally Activated Rupture Processes, Physical Review Letters 94 (3) (2005) 038501+.

[40] V. Filimonov, D. Sornette, Self-excited multifractal dynamics, Europhysics Letters 94 (4) (2011) 46003.

[41] A. Saichev, A. Helmstetter, D. Sornette, Anomalous Scaling of Offspring and Generation Numbers in Branching Processes, Pure and Applied Geophysics 162 (2005) 1113–1134.

[42] D. Harte, PtProcess: An R package for modelling marked point processes indexed by time,

Journal of Statistical Software 35 (8) (2010) 1–32.

[43] R. B. Schnabel, J. E. Koonatz, B. E. Weiss, A modular system of algorithms for unconstrained minimization, ACM Transactions on Mathematical Software 11 (4) (1986) 419–440.

[44] J. E. Dennis, R. B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Vol. 16 of Classics in Applied Mathematics, Society for Industrial Mathematics, 1987.

[45] F. Papangelou, Integrability of Expected Increments of Point Processes and a Related Random Change of Scale, Transactions of the American Mathematical Society 165 (1972) 483–506.

[46] J.-M. Zakoian, C. Francq, GARCH Models: Structure, Statistical Inference and Financial Applications, Wiley-Blackwell, Oxford, 2010.

[47] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control 19 (6) (1974) 716–723.

[48] P. Guttorp, T. L. Thorarinsdottir, Bayesian Inference for Non-Markovian Point Processes, in: E. Porcu, J. M. Montero, M. Schlather (Eds.), Advances and Challenges in Space-time Modelling of Natural Events, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 79–102.

[49] E. F. Fama, Efficient capital markets: II, Journal of Finance 46 (5) (1991) 1575–1617.

[50] E. F. Fama, Efficient Capital Markets: A Review of Theory and Empirical Work, The Journal of Finance 25 (2) (1970) 383–417.

[51] R. J. Shiller, Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?, The American Economic Review 71 (3) (1981) 421–436.

[52] D. M. Cutler, J. M. Poterba, L. H. Summers, What moves stock prices?, Journal of Portfolio Management 15 (3) (1987) 4–12.

[53] D. Sornette, Y. Malevergne, J.-F. Muzy, What causes crashes?, Risk 16 (2) (2003) 67–71.

[54] A. Joulin, A. Lefevre, D. Grunberg, J.-P. Bouchaud, Stock price jumps: news and volume play a minor role, Wilmott Magazine Sep/Oct (2008) 46.

[55] R. F. Engle, The Econometrics of Ultra-High-Frequency Data, Econometrica: Journal of the Econometric Society 68 (1) (2000) 1–22.

[56] M. Pacurar, Autoregressive Conditional Duration (ACD) Models in Finance: A Survey of the Theoretical and Empirical Literature, Journal of Economic Surveys 22 (4) (2008) 711–751.

[57] R. F. Engle, J. R. Russell, Analysis of High Frequency Financial Data, in: Handbook of Financial Econometrics, North Holland, 2009, pp. 383–426.

[58] G. Soros, The Alchemy of Finance: Reading the Mind of the Market, John Wiley & Sons, NY, 1987.

[59] J. D. Hamilton, Time Series Analysis, Princeton University Press, 1994.

[60] T. Mikosch, J.-P. Kreiß, R. A. Davis, T. G. Andersen (Eds.), Handbook of Financial Time Series, 1st Edition, Springer, 2009.

[61] P. A. W. Lewis, G. S. Shedler, Simulation of nonhomogeneous poisson processes by thinning, Naval Research Logistics Quarterly 26 (3) (1979) 403–413.

[62] Y. Ogata, On Lewis' simulation method for point processes, IEEE Transactions on Information Theory 27 (1) (1981) 23–31.

[63] J. Møller, J. G. Rasmussen, Perfect simulation of Hawkes processes, Advances in applied probability 37 (3) (2005) 629–646.

[64] J. Møller, J. G. Rasmussen, Approximate Simulation of Hawkes Processes, Methodology and Computing in Applied Probability 8 (1) (2006) 53–64.

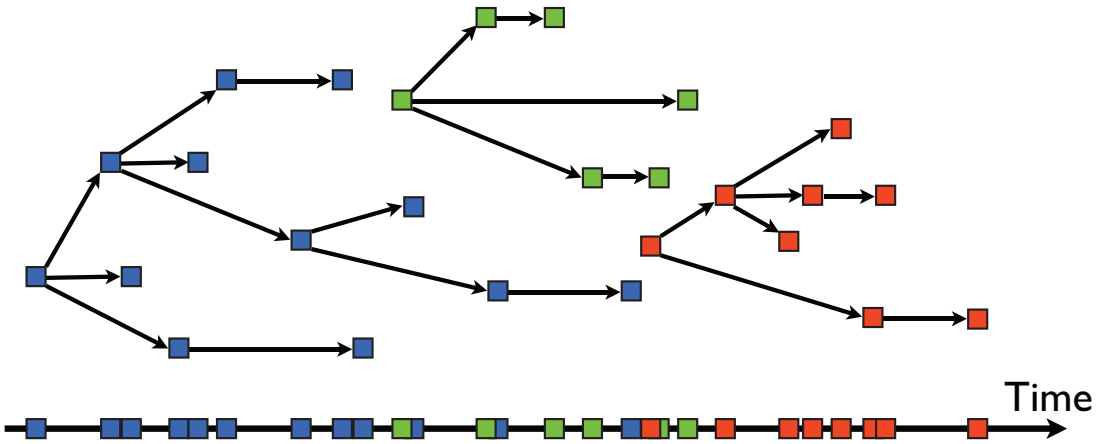

FIG. 1: (Color online) Illustration of the branching structure of the Hawkes process (top) and events on the time axis (bottom). Different colors of markers correspond to different clusters. This figure corresponds to a branching ratio $n = 0.88$.

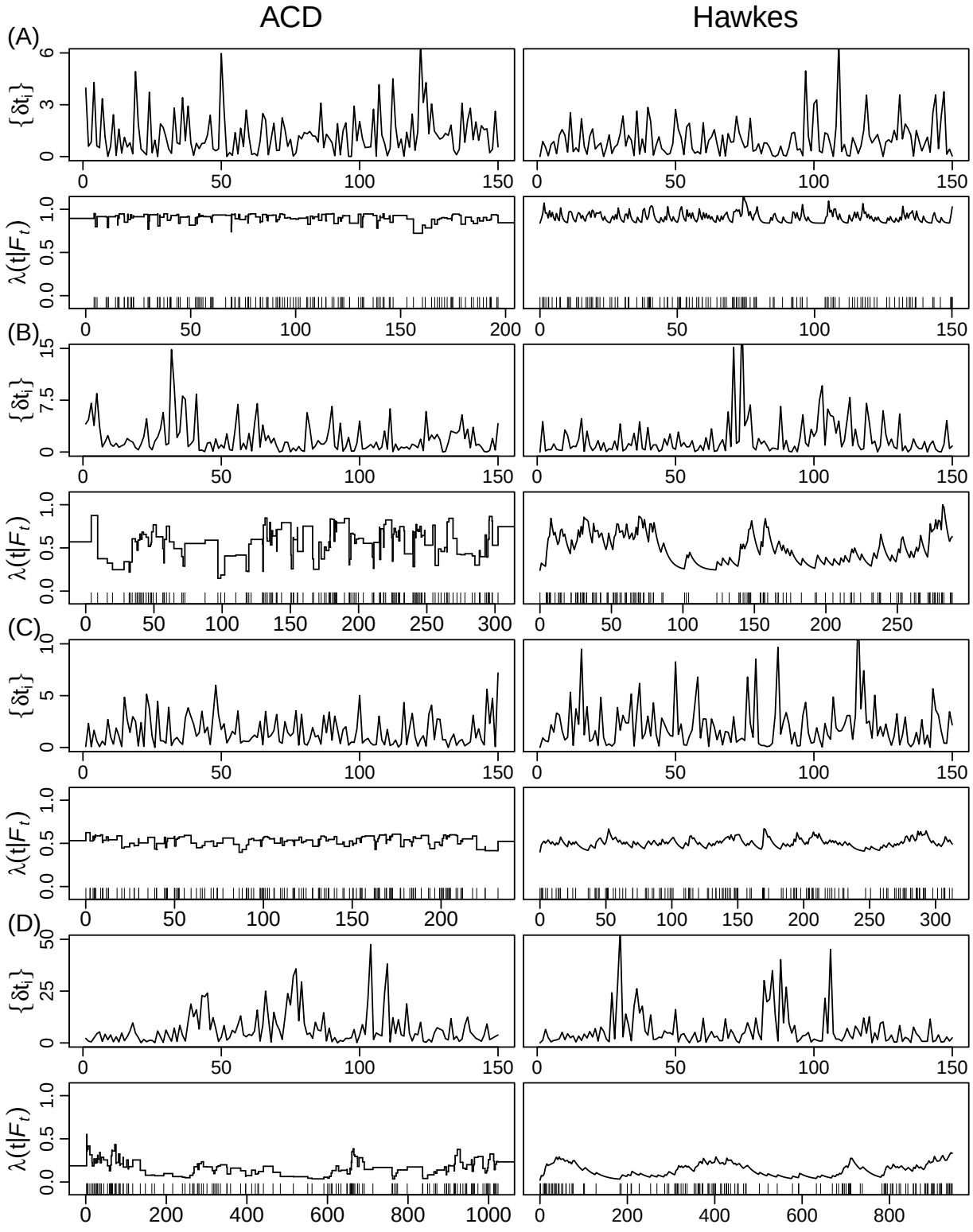

FIG. 2: Realizations of the durations and conditional intensities of the ACD(1,1) process (left column), and Hawkes process (right column) simulated with parameters obtained by calibrating to the realization of the ACD process. Parameters of the ACD process $\theta_{ACD} = (\omega, \alpha, \beta)$ and estimated parameters of the Hawkes model $\hat{\theta}_H = (\hat{\mu}, \hat{\eta}, \hat{\tau})$ are the following: (A) $\theta_{ACD} = (1, 0.05, 0.05)$, $\hat{\theta}_H = (0.84, 0.07, 4.3)$, (B) $\theta_{ACD} = (1, 0.38, 0.13)$, $\hat{\theta}_H = (0.24, 0.52, 5.6)$, (C) $\theta_{ACD} = (1, 0.13, 0.38)$,

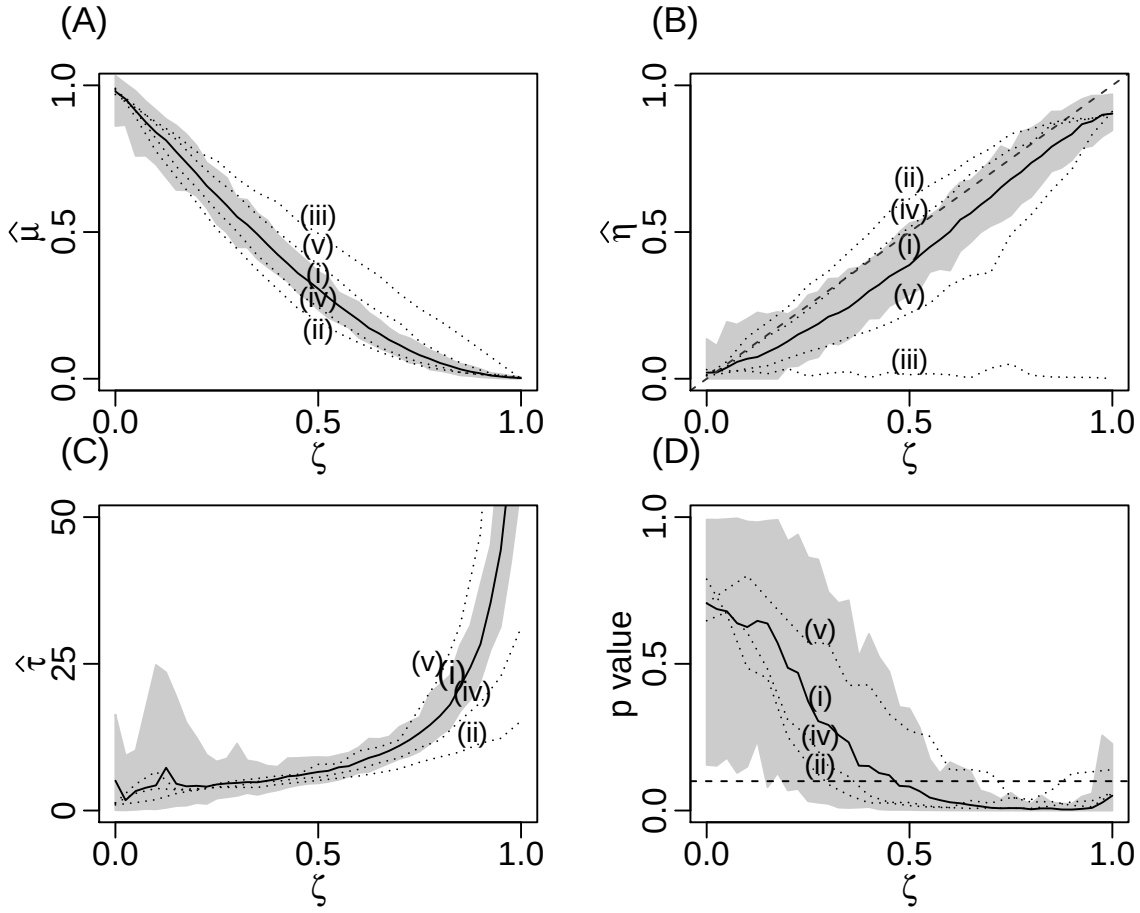

FIG. 3: Results of the calibration of the Hawkes model on the ACD(1,1) realizations. Estimated (A) background intensity $\hat{\mu}$, (B) branching ratio $\hat{\eta}$, and (C) characteristic time of the kernel $\hat{\tau}$. Panel (D) shows the p-value from the goodness-of-fit test, where the dashed line indicates the the 10% level. (i) $\alpha = \beta$ $(= \zeta/2)$, (ii) $\beta = 0$ $(\alpha = \zeta)$, (iii) $\alpha = 0$ $(\beta = \zeta)$, (iv) $\alpha = 3\beta$ $(= 3\zeta/4)$ and (v) $\beta = 3\alpha$ $(= 3\zeta/4)$. The black line corresponds to the mean p-value for case (i) $(\alpha = \beta)$, the shaded area to the 95% quantile range for case (i), and the dotted lines depict mean p-values for cases (ii) $\beta = 0$, (iii) $\alpha = 0$, (iv) $\alpha = 3\beta$ and (v) $\beta = 3\alpha$.

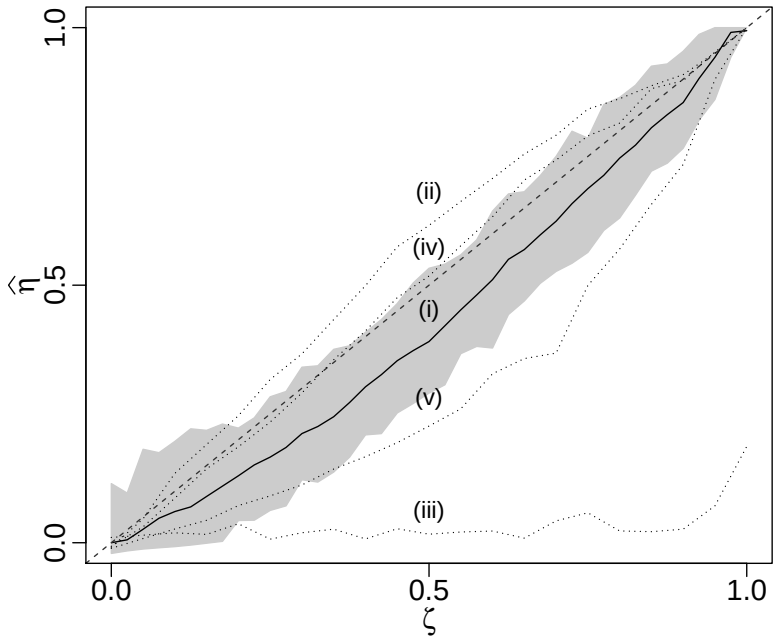

FIG. 4: The estimated branching ratio $\hat{\eta}$ of the Hawkes model estimated on ACD(1,1) realizations with correction for the finite sample estimation bias determined in the Appendix.

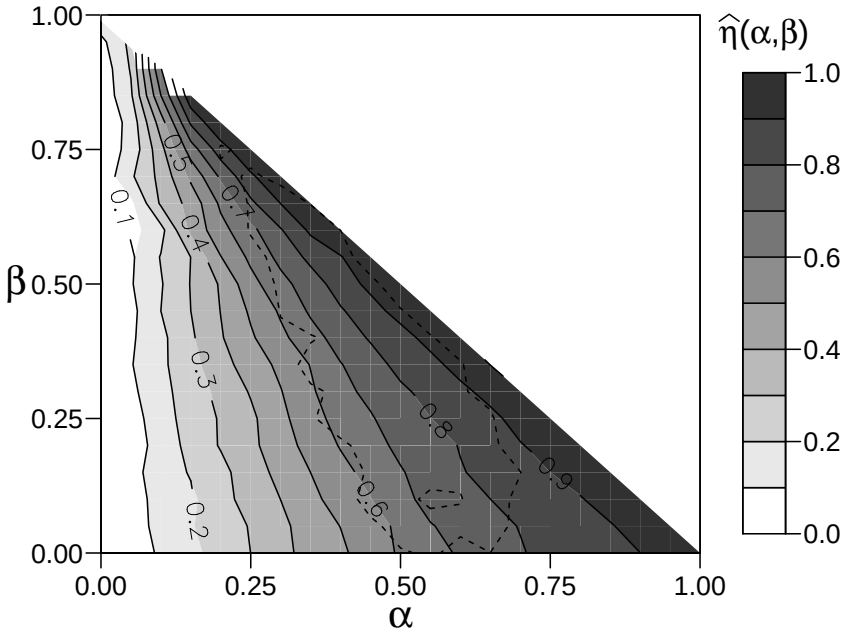

FIG. 5: Contour plot of the estimated branching ratio $\hat{\eta}(\alpha, \beta)$ of the Hawkes model calibrated to ACD(1,1) realizations for a grid of values α and β with $\alpha + \beta \leq 1$, corrected for the finite sample estimation bias determined in the Appendix. The dashed line delineates the region where the goodness-of-fit tests rejects the null hypothesis (see text).

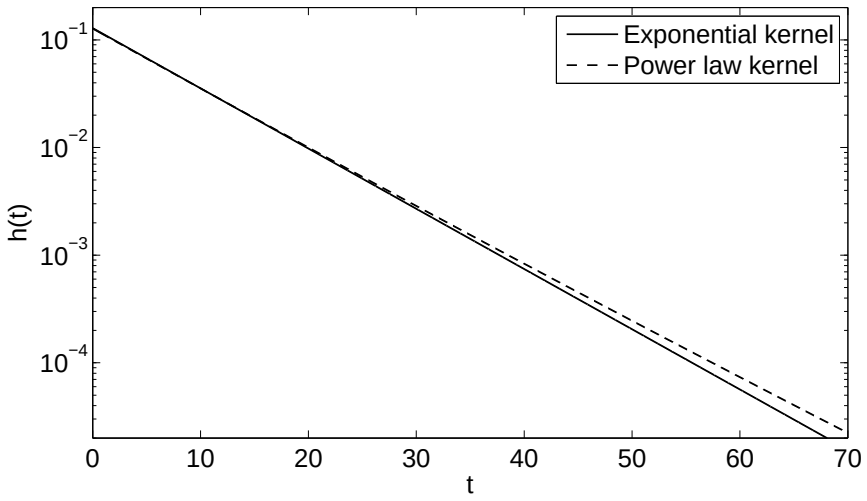

FIG. 6: Comparison of the exponential kernel with parameter $\tau = 7.9$ (solid line) with the power law kernel with parameters $\varphi = 105.17$ and $c = 816.41$ (dashed line).

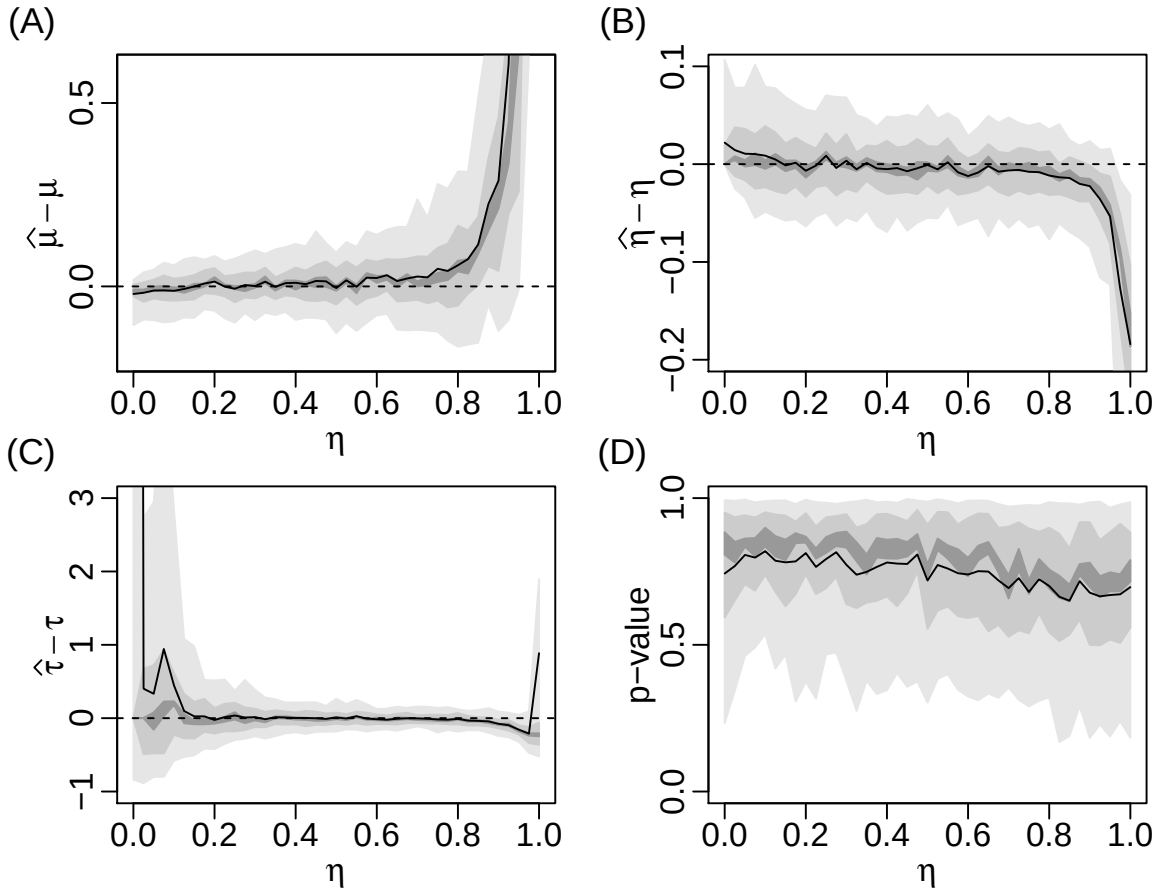

FIG. 7: Illustrations of the finite sample bias and variance of the maximum likelihood estimator [17] of the parameters of the Hawkes process calibrated on time series generated by the Hawkes process itself (no model error). Panel (A): difference between the estimates of the background intensity $\hat{\mu}$ and the true value μ used for the generation of the time series; Panel (B): difference between the estimates of the branching ratio $\hat{\eta}$ and the true value η; Panel (C): difference between the estimates of the characteristic time of the kernel $\hat{\tau}$ and the true value τ. Panel (D) shows the p-value of the Kolmogorov Smirnov test for standard uniformity of the transformed durations of the residual process [7]. In all panels, the black lines correspond to the mean, and the shaded areas to 90%, 50%, and 10% quantile ranges.