9

Time Clustering and Hawkes Processes

In economics, there can never be a "theory of everything". But I believe each attempt comes closer to a proper understanding of how markets behave.

(Benoît B. Mandelbrot)

When studying order flows in real markets, two striking phenomena are readily apparent:

- Market activity follows clear **intra-day patterns** that are related to predictable events, such as the start of a trading day, the end of a trading day, the opening of another market, the slowdown around lunch time, and scheduled macroeconomic news announcements. These intra-day patterns are clearly visible in Section 4.2. Such patterns induce strong 24-hour periodicities in several important market properties such as volatility (see, e.g., Figure 2.5). Other patterns also exist at the weekly, monthly, and yearly levels.
- Even after accounting for these intra-day patterns, market events still strongly cluster in time. Some periods have extremely high levels of market activity, with seemingly random durations, such that the level of market activity is "bursty" or "intermittent". In fact, financial markets are one among many other examples of **intermittent processes**, like neuronal activity or Barkhausen noise in disordered magnets (see references in Section 9.7). A well-known case is seismic activity, where one earthquake triggers aftershocks or even other major earthquakes elsewhere on the planet.

Throughout the previous four chapters, one of the key assumptions was that order flow can be described as a Poisson process with constant rate parameters. Although this assumption provides a convenient framework for building simple models, it clearly does not reflect these two important empirical facts.

In this chapter, we consider a class of models called *Hawkes processes*, which capture these properties via two key mechanisms. First, to account for exogenous

This chapter is not essential to the main story of the book.

intra-day patterns, they allow the Poissonian rate parameters to be explicitly time-dependent. Second, they incorporate a "self-exciting" feature, such that their local rate (or intensity) depends on the history of the process itself. As we illustrate throughout the chapter, these two core components make Hawkes processes powerful tools that strongly outperform homogeneous Poisson processes for replicating the true order-flow dynamics observed in real markets.

9.1 Point Processes

Throughout this chapter, our main object of study will be stochastic processes that create instantaneous temporal events. For $i \in \mathbb{N}$, we write t_i to denote the time at which the i^{th} event occurs. These events could be, for example, market order arrivals, changes in the mid-price, or even any event that changes the state of the LOB. The set of arrival times defines a **point process** (PP).

For a given point process $(t_i)_{i\in\mathbb{N}}$, the **inter-arrival times** are defined as

$$\delta_i = (t_{i+1} - t_i), \; i = 0, 1, 2, \ldots. \tag{9.1}$$

Given a PP $(t_i)_{i\in\mathbb{N}}$, the associated **counting process**

$$N(t) := \sum_{i\in\mathbb{N}} \mathbf{1}_{t_i \le t} \tag{9.2}$$

counts the number of arrivals that have occurred up to (and including) time t, where $\mathbf{1}_{t_i \le t}$ denotes the indicator function

$$\mathbf{1}_{t_i \le t} := \begin{cases} 1, & \text{if } t_i \le t, \\ 0, & \text{otherwise.} \end{cases} \tag{9.3}$$

Intuitively, the counting process $N(t)$ is an increasing step function with a unit-sized jump discontinuity at each time t_i.

For an infinitesimal time increment $\mathrm{d}t$, we define the **counting increment**

$$\mathrm{d}N(t) := \begin{cases} 1, & \text{if there exists an } i \text{ such that } t \le t_i \le t + \mathrm{d}t, \\ 0, & \text{otherwise.} \end{cases} \tag{9.4}$$

9.1.1 Homogeneous Poisson Processes

A **homogeneous Poisson process** is a PP in which events occur independently of each other and with a fixed intensity φ per unit time, such that the probability that $\mathrm{d}N(t) = 1$ in an infinitesimal time interval $(t, t + \mathrm{d}t)$ is equal to $\varphi \mathrm{d}t$.

An important property of a homogeneous Poisson process is that the inter-arrival times δ_i are independent random variables with an exponential distribution:

$$\mathbb{P}[\delta_i > \tau] = \mathbb{P}[N(t_i + \tau) - N(t_i) = 0] = e^{-\varphi\tau}, \qquad i = 0, 1, 2, \ldots.$$

Because the events are independent, the number of events that occur in an interval $(t, t+\tau)$ obeys a Poisson distribution:

$$\mathbb{P}[N(t+\tau) - N(t) = n] = \frac{(\varphi\tau)^n}{n!} e^{-\varphi\tau}. \tag{9.5}$$

By standard results for the Poisson distribution, the mean and variance of the number of events that occur in an interval $(t, t+\tau)$ are both equal to $\varphi\tau$:

$$\mathbb{E}[N(t+\tau) - N(t)] = \varphi\tau, \tag{9.6}$$

$$\mathbb{V}[(N(t+\tau) - N(t)] = \varphi\tau. \tag{9.7}$$

9.1.2 The Clustering Ratio

When studying PPs, the **clustering ratio**

$$r(\tau) = \frac{\mathbb{V}[(N(t+\tau) - N(t)]}{\mathbb{E}[(N(t+\tau) - N(t)]} \tag{9.8}$$

is a useful tool for understanding the temporal clustering of events:

- The value $r = 1$ (which, by Equations (9.6) and (9.7), is the case for a homogeneous Poisson process) indicates that events do not cluster in time.
- Values $r > 1$ indicate that arrivals attract each other (so that many arrivals can occur in the same interval), and therefore indicate the presence of clustering.
- Values $r < 1$ indicate that arrivals repel each other, and therefore indicate the presence of inhibition (see Section 9.5).

As we will see below, events in financial time series are characterised by values of r larger (and often much larger) than 1.

9.1.3 Inhomogeneous Poisson Processes

An **inhomogeneous Poisson process** is a Poisson PP with a time-dependent intensity $\varphi(t)$, which describes the instantaneous arrival rate of events at time t. The intensity is defined as

$$\varphi(t) = \mathbb{E}\left[\frac{\mathrm{d}N(t)}{dt}\right] = \lim_{\epsilon\downarrow 0} \mathbb{E}\left[\frac{N(t+\epsilon) - N(t)}{\epsilon}\right]. \tag{9.9}$$

One simple application of an inhomogeneous Poisson process is to choose $\varphi(t)$ as a deterministic function that tracks the intra-day pattern of activity in financial markets. Another possible application is to consider the intensity $\varphi(t)$ to itself be a random variable, with a certain mean φ_0 and covariance $\mathbb{V}[\varphi(t)\varphi(t')]: = C_\varphi(t,t')$. This "doubly stochastic"[1] point process is called a **Cox process**. Note

[1] The random nature of the intensity adds to the Poisson randomness.

that whenever $C_\varphi(t,t') > 0$ for $t' \neq t$, it follows that the clustering ratio satisfies $r > 1$, which indicates clustering.

9.2 Hawkes Processes

By allowing its rate parameter to vary as a function of time, an inhomogeneous Poisson process can address the first of the two empirical properties of order-flow series that we listed at the beginning of this chapter, namely predictable patterns. However, to also account for the second property (i.e. self-excitation) will require φ to depend not only on t, but on the past realisations of the process itself. This idea forms the basis of the Hawkes process.

9.2.1 Motivation and Definition

The inspiration for Hawkes processes originates not from financial markets, but from earthquake dynamics, where events seem to *trigger* other events, or replicas. This causes a point process of arrival times of earthquakes to exhibit a self-exciting structure, in which earthquakes are more likely to occur if other earthquakes have also occurred recently. The same phenomenon also occurs in financial markets. For example, trades occur frequently when many other trades have occurred recently, and occur infrequently when few other trades have occurred recently.

Hawkes processes are a specific family of inhomogeneous Poisson PPs where the temporal variation of $\varphi(t)$ has not only an **exogenous component** (such as the processes that we described in Section 9.1.3), but also an **endogenous component** that depends on the recent arrivals of the PP itself. More precisely, a **Hawkes process** is a point process whose arrival intensity is given by

$$\varphi(t) = \varphi_0(t) + \int_{-\infty}^{t} \mathrm{d}N(u)\Phi(t-u), \tag{9.10}$$

where $\varphi_0(t)$ is a deterministic base intensity and $\Phi(t) \geq 0$ is a non-negative **influence kernel** that describes how past events influence the current intensity. Often $\Phi(t)$ is chosen to be a strictly decreasing function of t, meaning that the influence of past events fades away with time.

Importantly, the integral in Equation (9.10) is calculated with respect to the counting process $N(u)$. The purpose of doing so is to sum the contributions to $\varphi(t)$ caused by all previous arrivals – which are precisely what $N(u)$ counts. Since $\mathrm{d}N(u)$ is equal to 1 if and only if one of the event-times t_i falls in the interval $[u, u+\mathrm{d}u]$, one can rewrite Equation (9.10) by replacing this integral with a sum that runs over the realised arrival times of the point process:

$$\varphi(t) = \varphi_0(t) + \sum_{t_i<t} \Phi(t-t_i). \tag{9.11}$$

As illustrated by Equation (9.11), the intensity function of a Hawkes process consists of two parts:

- The first part (i.e. the $\varphi_0(t)$ term) is deterministic, and describes the exogenous contribution to the rate dynamics. Its temporal variation might represent, for example, the intra-day periodicity of the activity.
- The second part (i.e. the $\sum_{t_i<t} \Phi(t - t_i)$ term) describes the endogenous contribution to the rate dynamics, which determines how previous arrivals affect the present arrival intensity. Because the second part of the intensity function depends on the history of the stochastic point process, the intensity function of a Hawkes process is itself a stochastic process.

Hawkes processes are widely used to model high-frequency market microstructure events in calendar time. As we will see throughout the remainder of this chapter, they offer several attractive benefits, including simplicity, flexibility, their ability to account for a wide range of different event types in a multivariate setting, their ability to account for non-stationarities through the exogenous $\varphi_0(t)$ term, and their appealing mathematical and statistical properties. Their parameters also have a simple and useful interpretation, and they thereby lead to a concise description of many complex aspects of market microstructure.

9.2.2 Basic Properties of Hawkes Processes

To simplify our discussion, we first consider some basic properties of Hawkes processes in the case where the exogenous intensity φ_0 is time-independent. We also assume that the process starts at $t = 0$ with no past (i.e. $N(0) = 0$). In this framework, it is possible to derive several interesting results about the statistical properties of the process.

First-Order Statistics

We first consider the mean intensity $\bar{\varphi}$ of the process, defined as the long-term growth rate of the counting process:

$$\bar{\varphi} := \lim_{t\to\infty} \frac{N(t)}{t}.$$

Provided that the process reaches a **stationary state** such that this quantity exists, $\bar{\varphi}$ must be equal to the long-term mean value of the intensity, after the influence of the initial state of the point process has died out:

$$\bar{\varphi} = \lim_{t\to\infty} \mathbb{E}[\varphi(t)].$$

Since $\mathbb{E}[\mathrm{d}N(u)] = \varphi(u)\mathrm{d}u$, one finds from Equation (9.10) that $\bar{\varphi}$ must obey the self-consistent equation

$$\bar{\varphi} = \varphi_0 + \bar{\varphi} \int_0^\infty \mathrm{d}u\, \Phi(u),$$

which only has a non-negative solution when the norm[2]

$$g := \int_0^\infty \mathrm{d}u\, \Phi(u) \tag{9.12}$$

of the kernel satisfies $g < 1$. Note that this condition requires $\Phi(t)$ to decrease sufficiently quickly for the integral to converge, and to be small enough for the process to reach a stationary state.

Let $g_c = 1$ denote the critical value for g. When $g < g_c$, the mean intensity of the Hawkes process is given by

$$\bar{\varphi} = \frac{\varphi_0}{1-g} \geq \varphi_0, \tag{9.13}$$

which indicates that the Hawkes self-exciting mechanism causes the mean intensity to increase. In the extreme case where $g \to g_c = 1$, this feedback is so strong that $\bar{\varphi}$ diverges. When $g > 1$, the process explodes, in the sense that $N(t)$ grows faster and faster with increasing t, so that no stationary state can ever be reached and the above mathematical formulas are meaningless. In the limit $g = 0$, there is no feedback, so $\bar{\varphi} = \varphi_0$ and we recover the homogeneous PP.

Second-Order Statistics

For a stationary Hawkes process with $g < 1$, the function

$$c(\tau) := \frac{1}{\bar{\varphi}^2} \mathrm{Cov}\left[\frac{\mathrm{d}N(t)}{dt}, \frac{\mathrm{d}N(t+\tau)}{dt}\right]; \qquad \tau > 0$$

describes the rescaled covariance of the arrival intensity. Note that since $\mathrm{d}N(t)^2 = \mathrm{d}N(t)$, there is a singular $\bar{\varphi}\delta(\tau)$ contribution to the equal-time covariance.

Observe that

$$\mathbb{E}[\mathrm{d}N(t)\mathrm{d}N(u)] = \mathbb{E}[\mathrm{d}N(t)|\mathrm{d}N(u) = 1]\mathbb{E}[\mathrm{d}N(u)].$$

Recalling also that $\mathbb{E}[\mathrm{d}N(u)] = \bar{\varphi}\mathrm{d}u$, it follows that the function $c(\cdot)$ is related to conditional expectation of the intensity function, as:

$$\mathbb{E}[\mathrm{d}N(t)|\mathrm{d}N(u) = 1] := \bar{\varphi}\,(1 + c(t-u))\,\mathrm{d}t + \delta(t-u). \tag{9.14}$$

Now, from the definition of the Hawkes process in Equation (9.10), one also has, for $t > u$:

$$\mathbb{E}[\mathrm{d}N(t)|\mathrm{d}N(u) = 1] = \left[\varphi_0 + \int_0^\infty \mathrm{d}v\, \Phi(v)\mathbb{E}[\mathrm{d}N(t-v)|\mathrm{d}N(u) = 1]\right]\mathrm{d}t.$$

[2] This quantity is also sometimes called the branching ratio (see Section 9.2.6).

Using Equation (9.14), the second term on the right-hand side can be transformed into

$$\int_0^{\infty} \mathrm{d}v\, \Phi(v)\mathbb{E}[\mathrm{d}N(t-v)|\mathrm{d}N(u)=1] = \bar{\varphi}g + \Phi(t-u) + \bar{\varphi}\int_0^{\infty} \mathrm{d}v\Phi(v)c(t-v-u).$$

By comparing the two expressions for $\mathbb{E}[\mathrm{d}N(t)|\mathrm{d}N(u)=1]$ and noting that $\varphi_0 + g\bar{\varphi} = \bar{\varphi}$, we finally arrive at the **Yule–Walker equation**, for $\tau > 0$:

$$c(\tau) = \frac{1}{\bar{\varphi}}\Phi(\tau) + \int_0^{\infty} \mathrm{d}u\, \Phi(u)c(\tau-u); \qquad c(-\tau) = c(\tau). \tag{9.15}$$

Introducing

$$\widehat{c}(z) = \int_{-\infty}^{\infty} \mathrm{d}u c(u) e^{-zu}, \quad \widehat{\Phi}(z) = \int_0^{\infty} \mathrm{d}u\, \Phi(u) e^{-zu},$$

this equation turns out to be equivalent to[3]

$$1 + \bar{\varphi}\widehat{c}(z) = \frac{1}{(1-\widehat{\Phi}(z))(1-\widehat{\Phi}(-z))}. \tag{9.16}$$

Equation (9.15) thus provides a way to compute the covariance of the arrival intensity $c(\cdot)$, given the Hawkes kernel Φ. Conversely, it also provides a way to reconstruct the kernel of the underlying Hawkes process from an empirical determination of the covariance c. In other words, the rescaled covariance of the intensity $c(\tau)$ fully characterises the Hawkes kernel $\Phi(\tau)$, while the average intensity $\bar{\varphi}$ allows one to infer the bare intensity φ_0. Much like the Gaussian process, the Hawkes process is entirely determined by its first- and second-order statistics.[4]

9.2.3 Some Useful Results

By noting that $\widehat{\Phi}(0) = g$, we obtain the identity

$$\bar{\varphi}\widehat{c}(0) := \bar{\varphi}\int_{-\infty}^{\infty} \mathrm{d}u\, c(u) = \frac{g(2-g)}{(1-g)^2}, \tag{9.17}$$

which is valid for any kernel shape. Importantly, this identity tells us that the integrated rescaled covariance is only determined by the branching ratio g of the Hawkes process.

[3] Deriving this result is non-trivial due to the infinite upper limit of the integral on the right-hand side of Equation (9.15). For a derivation, see Bacry, E., & Muzy, J.-F. (2016). First- and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory*, 62(4), 2184–2202.

[4] Note that by the same token, it is hard to distinguish correlation and causality, since two-point correlation functions are always invariant under time reversal. Therefore, fitting a point process using a Hawkes model in no way proves the existence of genuinely causal self-exciting effects. For an extended discussion, see Blanc, P., Donier, J., & Bouchaud, J. P. (2016). Quadratic Hawkes processes for financial prices. *Quantitative Finance*, 17, 1–18.

From the covariance of the arrival intensity, one can compute the variance of the number of events $N(t+\tau) - N(t)$ in an interval of size $\tau > 0$:

$$\mathbb{V}[N(t+\tau) - N(t)] = \int_t^{t+\tau} \int_t^{t+\tau} \mathbb{E}[\mathrm{d}N(u)\mathrm{d}N(u')]$$
$$= \bar{\varphi}\tau + 2\bar{\varphi}^2 \int_0^\tau \mathrm{d}u\,(\tau - u)c(u), \tag{9.18}$$

where the first term comes from the singular, equal-time contribution to c, and the second term is zero for a homogeneous Poisson process, for which $\Phi = c = 0$. It also follows from Equation (9.18) that the clustering ratio (9.8) is:

$$r(\tau) = 1 + 2\bar{\varphi} \int_0^\tau \mathrm{d}u \left(1 - \frac{u}{\tau}\right) c(u). \tag{9.19}$$

Note that for a pure Poisson process, $c(u) = 0$, and we recover the standard result $r = 1$. For a positive Hawkes kernel Φ, the value of $c(u)$ becomes positive as well, which leads to clustering (i.e. $r > 1$).

If we assume that $c(u)$ decreases faster than u^{-2} for large u, so that $\int_0^\infty \mathrm{d}u\, uc(u)$ is a finite integral, then Equation (9.19) can be simplified at large τ, as

$$r(\tau) \approx_{\tau\to\infty} 1 + 2\bar{\varphi} \int_0^\infty \mathrm{d}u\, c(u) = 1 + \bar{\varphi} \int_{-\infty}^\infty \mathrm{d}u\, c(u) = \frac{1}{(1-g)^2}, \tag{9.20}$$

where we have used the identity (9.17). This useful result can be used to estimate the feedback parameter g from empirical data, without having to specify the shape of the Hawkes kernel Φ.

9.2.4 Example I: The Exponential Kernel

To illustrate the application and interpretation of some of these results, we first consider the case of an **exponential kernel**

$$\Phi(t \geq 0) = g\omega e^{-\omega t}, \qquad t \geq 0, \tag{9.21}$$

where ω^{-1} sets the time scale beyond which past events can be regarded not to influence present intensity, and g is the norm of the kernel. Using Equation (9.15) in Laplace space, one readily derives:

$$c(t > 0) = \frac{\omega g(2-g)}{2\varphi_0} \omega e^{-(1-g)\omega t}. \tag{9.22}$$

This result is very interesting, because it tells us that the Hawkes feedback not only increases the mean intensity of the process (as intended), but also increases the memory time of the process. While the direct self-exciting mechanism dies out in a time ω^{-1}, the induced extra activity survives until a time $\omega^{-1}/(1-g)$, which diverges as $g \to 1$. We present an interpretation of this effect in terms of population dynamics, or branching process, in Section 9.2.6.

The clustering ratio r can also be calculated exactly, using Equations (9.19) and (9.22). One finds that

$$r = \frac{1}{(1-g)^2} - \frac{g(2-g)}{(1-g)^3\omega\tau}(1 - e^{-(1-g)\omega\tau}), \tag{9.23}$$

which rises from the Poisson value $r = 1$ when $\tau \ll \omega^{-1}$ to the asymptotic value $r = (1-g)^{-2}$ when $\tau \gg \omega^{-1}$.

Note that it is straightforward to simulate a Hawkes process with an exponential kernel. The trick is to simulate the process in event-time by calculating the (calendar) arrival times. Assume n events have already arrived at times $t_1, t_2, \cdots, t_n$. We need to determine the probability density of the arrival time of the $(n+1)^{\text{th}}$ event. The cumulative law of the inter-arrival times δ_n is given by

$$\mathbb{P}[\delta_n \leq \tau] = 1 - \exp\left[-\varphi_0\tau - g\,(1 - e^{-\omega\tau})\sum_{k\leq n} e^{-\omega(t_n - t_k)}\right]. \tag{9.24}$$

One can now generate the time series of arrival times iteratively from Equation (9.24) which can be achieved numerically by inverse-transform sampling.

9.2.5 Example II: The Near-Critical Power-Law Kernel

Consider now the case of a **power-law kernel**

$$\Phi(t \geq 0) = g\frac{\omega\beta}{(1+\omega t)^{1+\beta}}, \tag{9.25}$$

where the parameter $\beta > 0$ ensures that the integral defining the norm of Φ is convergent. In the limit $z \to 0$, the behaviour of $\widehat{\Phi}(z)$ depends on the value of β. For $\beta < 1$, asymptotic analysis leads to

$$\widehat{\Phi}(z) \approx g - g\Gamma[1-\beta](z/\omega)^{\beta}.$$

Inserting this behaviour into Equation (9.16) allows one to derive the behaviour of the correlation function $c(t)$. For $g \to 1$, the inversion of the Laplace transform gives a power-law behaviour for $c(t)$ at large times, as[5]

$$c(t) = c_\infty t^{-\gamma}; \qquad \gamma = 1 - 2\beta. \tag{9.26}$$

This result only makes sense if $\beta < 1/2$; otherwise, γ would be positive and $c(t)$ would grow without bound (which is clearly absurd). In fact, this is precisely

[5] See Appendix A.2. The prefactor identification gives $c_\infty^{-1} = 2g^2\omega^{-2\beta}\cos(\pi\beta)\Gamma^2[1-\beta]\Gamma[2\beta]$.

the result of Brémaud and Massoulié: the critical Hawkes process only exists for power-law kernels that decay with an exponent β in the range $0 < \beta < 1/2$, such that $\gamma = 1 - 2\beta$ satisfies $0 < \gamma < 1$. For $\beta > 1/2$ and $g = 1$, the Hawkes process never reaches a stationary state, or is trivially empty. Conversely, when the correlation function decays asymptotically as a power law with $\gamma < 1$, the Hawkes process is necessarily critical (i.e. $g = 1$).

When $\beta < 1/2$ and $g = 1 - \epsilon$ with $\epsilon \ll 1$, the correlation function $c(t)$ decays as $t^{-\gamma}$ up to a (long) crossover time $\tau^* \sim \epsilon^{-1/\beta}$, beyond which it decays much faster, such that $c(t)$ is integrable.

Finally, one can compute the large-τ behaviour of the clustering ratio r:

$$r(\tau) \sim \frac{2c_\infty}{(1-\gamma)(2-\gamma)} \tau^{1-\gamma}, \qquad (\tau \ll \tau^*),$$

which grows without bound when $\tau^* \to \infty$ (i.e. $g \to 1$).

9.2.6 Hawkes Processes and Population Dynamics

By regarding them as **branching processes** (also known as Watson–Galton models), Hawkes processes can also model population dynamics. In a branching process, exogenous "mother" (or "immigration") events occur at rate φ_0, and endogenously produce "child" (or "birth") events, which themselves also endogenously produce child events, and so on. A single mother event and all its child events are collectively called a *family*.

In such models, an important quantity is the expected number g of child events that any single (mother or child) event produces. This number g is called the **branching ratio**. If $g > 1$, then each event produces more than one child event on average, so the population explodes. If $g < 1$, then each event produces less than one child event on average, so the endogenous influence of any single event eventually dies out (although in a Hawkes process, this does not necessarily correspond to the whole process dying out, due to the exogenous immigration arrivals, which are driven by φ_0). The mean family size $\bar{S}$ is $1/(1-g)$.

The critical case $g = g_c = 1$ is extremely interesting, because for some special choices of power-law kernels Φ, the population can survive indefinitely and reach a stationary state even in the limit of no "ancestors", when $\varphi_0 \to 0$. In this case, the family sizes S are distributed as a power-law with a universal exponent $S^{-3/2}$ (i.e. independently of the precise shape of Φ). We will see below that this special case, first studied by Brémaud and Massoulié,[6] seems to corresponds to the behaviour observable in financial markets.

[6] See: Brémaud, P., & Massoulié, L. (2001). Hawkes branching point processes without ancestors. *Journal of Applied Probability*, 38(01), 122–135.

The analogy with population dynamics also allows us to understand why the memory time of the process increases as the expected number of child events increases: while mother events only have a certain fertility lifetime ω^{-1}, the lifetime of the whole lineage (which keeps the memory of the process) can be much longer – in fact, by a factor of the mean family size, $1/(1-g)$.

9.3 Empirical Calibration of Hawkes Processes

At the beginning of this chapter, we remarked that two features of real markets render homogeneous Poisson processes as unsuitable for modelling event arrival times. The first feature is the existence of seasonalities; the second feature is the existence of temporal clustering (even after the seasonalities have been accounted for). As we have discussed throughout the chapter, a Hawkes process (as defined in Equation (9.10)) can address both of these problems by incorporating seasonalities via its exogenous arrival rate $\varphi_0(t)$ and incorporating clustering via the Hawkes kernel. In this section, we explain how Hawkes processes can be calibrated on data. We first discuss how to account for the observed intra-day seasonalities, then we present a selection of methods for estimating the shape of the Hawkes kernel.

9.3.1 Addressing Intra-day Seasonalities

As we have discussed several times already, clear intra-day patterns influence activity levels in financial markets (see Figures 1.3 and 2.5 and Section 4.2). In the context of Hawkes processes, this feature can be addressed as follows. First, measure the intra-day pattern of activity $\bar{\varphi}(t)$ across many different trading days. Then, given $\bar{\varphi}(t)$, define **business time** as

$$\hat{t}(t) = \int_0^t \mathrm{d}u\, \bar{\varphi}(u).$$

In this re-parameterised version of calendar time, the speed of the clock is influenced by the average level of activity. This simple trick causes the activity profile to become flat in business time. We implement this technique (which may essentially be regarded as a pre-processing step) in the financial applications that we describe throughout the remainder of this chapter.

9.3.2 Estimating the Hawkes Kernel

We now turn to estimating the Hawkes kernel $\Phi(t)$. This kernel is characterised by two important features:

 https://doi.org/10.1017/9781316659335.014

- its time dependence (e.g. an exponential kernel with some well-defined decay time ω^{-1}); and
- its norm $g = \int_0^\infty \mathrm{d}u\,\Phi(u)$, which quantifies the level of self-reflexivity in the activity.

As we stated in Section 9.3.1, we assume that we have re-parameterised time to address intra-day seasonalities, and therefore that φ_0, when measured in business time, is constant. We first present the simplest way to elicit and quantify non-Poissonian effects, based on the clustering ratio r. Within the context of Hawkes processes, estimating r immediately leads to an estimate of the feedback intensity (or kernel norm) g, independently of the form of the kernel itself (see Equation 9.19). In the context of microstructural activity, we will see that g is not only positive (which indicates that clustering effects are present beyond the mere existence of intra-day patterns) but, quite interestingly, is not far from the critical value $g_c = 1$. We then proceed to the calibration of Hawkes processes using two traditional calibration processes: maximum likelihood estimation and the method of moments.

9.3.3 Direct Estimation of the Feedback Parameter

Recall from Equation (9.20) that for large τ,

$$r(\tau) \approx_{\tau\to\infty} \frac{1}{(1-g)^2}.$$

Empirically, one can extract M (possibly overlapping) intervals of size τ from a long time series of events, and compute from these M sub-samples the mean and variance of the number of events in such intervals.[7] The theoretical prediction of Equation (9.23) is very well-obeyed by the surrogate series generated using an exponential kernel, whereas for the financial data, the clustering ratio $r(\tau)$ does not seem to saturate even on windows of $\tau = 10^5$ seconds, but rather grows as a power-law of time. This means that as the time scale of observations increases, the effective feedback parameter $g_{\text{eff}}(\tau) := 1 - 1/\sqrt{r(\tau)}$ increases towards $g_c = 1$, and does so as a power-law:

$$g_{\text{eff}}(\tau) \cong 1 - \frac{A}{\tau^\beta},$$

where A is a constant and $\beta \cong 0.35$ (see the inset of Figure 9.1).

Returning to Equation (9.19), this power-law behaviour suggests that the correlation function $c(u)$ itself decays as a slow power-law, as we mentioned in

[7] See: Hardiman, S. J., & Bouchaud, J. P. (2014). Branching-ratio approximation for the self-exciting Hawkes process. *Physical Review E*, 90(6), 062807.

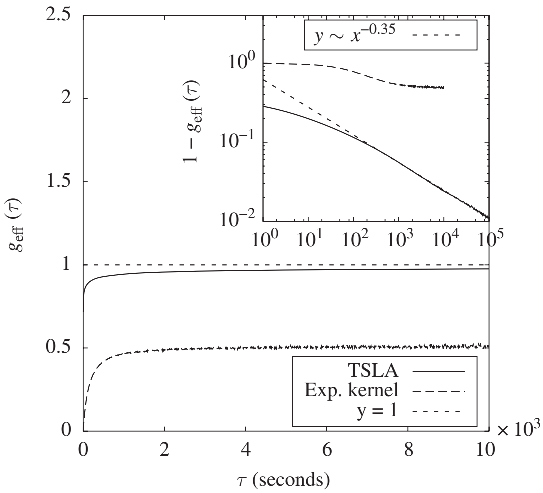

Figure 9.1. Estimation of the effective feedback parameter $g_{\text{eff}}(\tau)$ as a function of τ, measured in seconds, for (solid curve) mid-price changes of TSLA, and (dashed) a simulation of a Hawkes process using an exponential kernel with norm $g = 1/2$. We confirm our estimation procedure by observing that the simulated g_{eff} converges exponentially fast to its theoretical asymptotic value, $g_{\text{eff}}(\infty) \to g = 1/2$. The empirical time series exhibits criticality because $g_{\text{eff}}(\tau)$ converges to 1, (inset) as a power-law with exponent $\beta \cong 0.35$ (the dotted line shows a power law with exponent 0.35).

Section 2.2.4. More precisely, assuming that $c(u) \sim_{u\to\infty} c_\infty/u^\gamma$ with $\gamma < 1$, one obtains

$$r(\tau) = \frac{1}{(1 - g_{\text{eff}}(\tau))^2} \approx \frac{2c_\infty}{(1-\gamma)(2-\gamma)}\tau^{1-\gamma},$$

which indeed leads to $g_{\text{eff}}(\tau)$ approaching 1 as a power law, with $\beta = (1-\gamma)/2$ (leading to $\gamma \cong 0.3$). This is as expected for a critical Hawkes process with a power-law kernel – see Equation (9.26). We will return to this observation in Section 9.3.5.

9.3.4 Maximum-Likelihood Estimation

If we assume that the kernel function $\Phi(t)$ can be written in a parametric form, with parameter vector $\boldsymbol{\theta}$, then we can use standard **maximum-likelihood estimation** (MLE) to estimate $\boldsymbol{\theta}$ from a realisation of the point process. If N arrivals occur during the interval $[0, T)$, then the likelihood function $\mathbb{L}$ is given by the joint density of all the observed intervals $[t_0 = 0, t_1]$, $[t_1, t_2]$, … in $[0, T)$. First, note that the probability for an event to occur between t_i and $t_i + \mathrm{d}t$ but not at any intermediate times $t_{i-1} < t < t_i$ is given by:

$$\exp\left(-\int_{t_{i-1}}^{t_i} \mathrm{d}u\, \varphi(u|\boldsymbol{\theta})\right)\varphi(t_i|\boldsymbol{\theta})\mathrm{d}t.$$

The likelihood of the whole sequence of event times is then given by:

$$\mathbb{L}(t_1, t_2, \ldots, t_N|\boldsymbol{\theta}) = \left(\prod_{i=1}^{N} \varphi(t_i|\boldsymbol{\theta}) \exp\left(-\int_{t_{i-1}}^{t_i} \mathrm{d}u\, \varphi(u|\boldsymbol{\theta})\right)\right) \exp\left(-\int_{t_N}^{T} \mathrm{d}u\, \varphi(u|\boldsymbol{\theta})\right),$$
$$= \left(\prod_{i=1}^{N} \varphi(t_i|\boldsymbol{\theta})\right) \exp\left(-\int_{0}^{T} \mathrm{d}u\, \varphi(u|\boldsymbol{\theta})\right).$$

Taking logarithms of both sides yields the log likelihood

$$\ln \mathbb{L}(t_1, t_2, \ldots, t_N|\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln \varphi(t_i|\boldsymbol{\theta}) - \int_{0}^{T} \mathrm{d}u\, \varphi(u|\boldsymbol{\theta}). \tag{9.27}$$

The MLE $\boldsymbol{\theta}^*$ is known to have nice properties: it is both *consistent* (i.e. it converges to the true $\boldsymbol{\theta}$ when $T \to \infty$) and asymptotically normally distributed around the true $\boldsymbol{\theta}$ with a known covariance matrix, with all elements of order T^{-1}. However, because the likelihood itself depends on the set of arrival times $t_1, t_2, \ldots, t_N$, there does not exist a simple "one-size-fits-all" closed-form solution for the MLE of $\boldsymbol{\theta}$. In many cases, the simplest approach is to use a numerical optimisation algorithm to maximise Equation (9.11) directly. However, some specific choices of Φ lead to sufficiently simple parametric forms for Equation (9.11) to allow easy computation of this integral, and therefore of the MLE. One such example is the exponential kernel, which we now investigate in more detail.

The Exponential Kernel

The exponential Hawkes process, which is defined by the kernel in Equation (9.21), depends on three parameters: $\boldsymbol{\theta} = (\varphi_0, \omega, g)$. With this kernel, the Hawkes dynamics obeys the stochastic differential equation

$$\mathrm{d}\varphi(t) = -\omega(\varphi(t) - \varphi_0)\mathrm{d}t + g\omega \mathrm{d}N(t). \tag{9.28}$$

We can interpret this as $\varphi(t)$ relaxing exponentially towards its base intensity φ_0, but being jolted upwards by an amount $g\omega$ each time an event occurs.

This exponential form is popular in many practical applications, mainly due to its Markovian nature: the evolution of $\varphi(t)$ only depends on the present state of the system, because all of the past is encoded into the value of $\varphi(t)$ itself. Numerically, this enables fast updates of the quantities needed to compute the likelihood function. Indeed, one can write:

$$\ln \mathbb{L}(t_1, t_2, \ldots, t_N|\boldsymbol{\theta}) = -T\varphi_0 + \sum_{i=1}^{N} \left[\ln(\varphi_0 + g\omega Z_i) - g\left(e^{-\omega(T-t_i)} - 1\right)\right] \tag{9.29}$$

where we have introduced the notation

$$Z_i := \sum_{k=1}^{i-1} e^{-\omega(t_i - t_k)}.$$

Expressed in this way, it seems that evaluating the likelihood function is an $O(N^2)$ operation. However, by using another trick, this can be reduced further, to $O(N)$. The trick is to notice that Z_i can be computed as a simple recursion, since:

$$Z_i = e^{-\omega(t_i - t_{i-1})} \sum_{k=1}^{i-1} e^{-\omega(t_{i-1} - t_k)} = (1 + Z_{i-1})\, e^{-\omega(t_i - t_{i-1})}.$$

The same trick can also be used when the kernel is a weighted sum of K exponentials:

$$\Phi(t) = g \sum_{k=1}^{K} a_k \omega_k e^{-\omega_k t}, \qquad \sum_{k=1}^{K} a_k = 1. \tag{9.30}$$

To use the trick in this case, one has to introduce K different objects Z_i^k that all evolve according to their own recursion:

$$Z_i^k = e^{-\omega_k (t_i - t_{i-1})} \left(1 + Z_{i-1}^k\right).$$

In this way, the computation of $\mathbb{L}$ and its gradients with respect to the three parameters $\boldsymbol{\theta} = (\varphi_0, \omega, g)$ (which are needed to locate the MLE $\boldsymbol{\theta}^*$) can all be performed efficiently, even on long time series. This is important because (as we discuss in the next section) power-law kernels can be approximated as sums of a finite number of exponentials.

The Power-Law Kernel

The power-law Hawkes process defined in Equation (9.25),

$$\Phi(t \geq 0) = g \frac{\omega\beta}{(1 + \omega t)^{1+\beta}},$$

depends on four parameters: $\boldsymbol{\theta} = (\varphi_0, \omega, g, \beta)$. Interestingly, exponential kernels can be seen as a special limit of power-law kernels, when $\beta \to \infty$ and $\omega \to 0$ with a fixed value of the product $\omega\beta$. Therefore, exponential Hawkes processes are a sub-family of power-law Hawkes processes.

Conversely, a power-law kernel can be accurately reproduced by a sum of exponential kernels with different time scales and different weights.[8] In particular,

[8] On this point, see, e.g., Bochud, T., & Challet, D. (2007). Optimal approximations of power laws with exponentials: Application to volatility models with long memory. *Quantitative Finance*, 7(6), 585–589, and Hardiman, S. J., Bercot, N., & Bouchaud, J. P. (2013). Critical reflexivity in financial markets: A Hawkes process analysis. *European Physical Journal B*, 86, 442–447.

choosing a discrete set of time scales as a geometric series, i.e. $\omega_k = \omega/b^k$, $b > 1$, and weights as $a_k = \omega_k^{\beta}$, Equation (9.30) allows one to reproduce quite reasonably a power-law decaying kernel $\Phi(t) \sim t^{-1-\beta}$, at least for t such that $\omega^{-1} \ll t \ll b^K \omega^{-1}$.

When using the MLE method for estimating the parameters of power-law kernels on financial data, special care must be given to errors in the time-stamping of the events. Such errors can substantially affect the MLE procedure and lead to spurious solutions, due to artefacts arising from short-time dynamics. In any case, an MLE of Hawkes parameters applied to a wide variety of financial contracts suggests very high values of g^*, close to the critical value $g_c = 1$, and yields a power-law exponent β^* in the range 0.1–0.2 (see references in Section 9.7). This clearly excludes the possibility of an exponential kernel, which corresponds to $\beta \gg 1$. Interestingly, the characteristic time scale ω^{-1}, below which the power-law behaviour saturates, has decreased considerably over time, from several seconds in 1998 to milliseconds in 2012. This is clearly contemporaneous to the advent of execution algorithms and high-frequency trading, which vindicates the idea that any kind of feedback mechanism has accelerated during that period. Note, however, that in contrast with the exponential case, a power-law kernel plausibly suggests the existence of a wide spectrum of reaction times in the market, from milliseconds to days or longer.

There is, however, a problem with using this approach to calibrate a Hawkes process to empirical data. The MLE method implicitly over-focuses on the short-time behaviour of the kernel and is quite insensitive to the shape of the kernel $\Phi(t)$ at large lags. Indeed, since the power-law kernel gives much greater weight to short lags than to long lags, the local intensity φ is mostly determined by the recent events and much less by the far-away past. Therefore, the MLE procedure will mostly optimise the short-time behaviour of the kernel, and will be quite sensitive to the detailed shape of this kernel for $t \to 0$, as well as on the quality of the data for short times. This is why the value of the exponent β^* obtained using MLE is significantly smaller than the value of $\beta \cong 0.35$ obtained from the direct estimation method used in the previous section: the two methods zoom in on different regions of the time axis, as we elaborate on in the next paragraph.

9.3.5 The Method of Moments

Another popular tool for calibrating Hawkes process is the **method of moments**, which seeks to determine the parameters $\boldsymbol{\theta}$ such that the theoretical moments of specified observables match their empirically determined counterparts. Restricting this method to first- and second-order statistics (which is enough to completely determine a Hawkes process, see Section 9.2.2), this method amounts to fitting φ_0 and $\Phi(t)$ such that:

- the mean intensity $\bar{\varphi} = \varphi_0/(1-g)$ matches the empirically determined mean; and

- the predicted intensity correlation function $c(t)$ (using the Yule–Walker equation (9.15)) reproduces the empirically determined intensity correlation.

The first condition simply amounts to counting the number of events in sufficiently large intervals, but the second condition requires extracting the shape of the kernel $\Phi(t)$ using the empirical correlation and Equation (9.15), which is much less trivial.

In fact, seen as an expression for $\Phi(t)$ *given* $c(t)$, Equation (9.15) is called a **Wiener–Hopf equation**. Solving it requires complicated numerical methods, but once this is done, the kernel $\Phi(t)$ can be determined in a wide region of time scales[9] (provided $c(t)$ can be measured with sufficient accuracy). Again, implementing this method on empirical data leads to a power-law kernel with a norm close to 1. An alternative method consists in fitting the measured correlation function $c(t)$ with simple mathematical expressions, and inverting the Wiener–Hopf equation analytically.

The method of moments requires measuring the function $c(t)$ accurately. This can prove difficult, as it requires choosing an adequate bin size Δt to define the count increments ΔN. Interestingly, the clustering ratio $r(\tau)$ allows one to bypass this step, and provides simple and intuitive results. Plotting r as a function of τ gives direct access to the time dependence of $\int_0^\tau \mathrm{d}u\,(1 - u/\tau)c(u)$, and thus of $c(t)$ itself. As shown in Figure 9.1, $c(t)$ actually has not one but *two* power-law regimes: a short-time behaviour with $\beta \cong 0.2 - 0.25$ (compatible with MLE) and a long-time behaviour with $\beta \cong 0.35 - 0.4$. A similar crossover between the two regimes exists for, e.g., the S&P 500 futures contract, and occurs for τ of the order of several minutes.[10] Both regimes are characterised by exponents β in the Brémaud–Massoulié interval $0 < \beta < \frac{1}{2}$.

9.4 From Hawkes Processes to Price Statistics

In Section 2.2.4, we introduced a simple model in which changes of the mid-price $m(t)$ occur as a homogeneous Poisson process with rate φ, all have size ϑ (i.e. one tick), and occur upwards or downwards with equal probabilities. This model can be expressed concisely as:

$$\mathrm{d}m(t) = \varepsilon(t)\vartheta\,\mathrm{d}N(t), \tag{9.31}$$

where $N(t)$ is the counting process of mid-price changes and $\varepsilon(t) = \pm 1$ is an independent binary random variable. We now extend this model to the case where $\varphi(t)$ is time-dependent, such that $\mathrm{d}N(t)$ is given by a stationary Hawkes process with an influence kernel Φ.

[9] Note, however, that this method provides no guarantee that for a given $c(t)$, the kernel function Φ will be everywhere non-negative, as is required for the intensity φ.

[10] See: Hardiman et al. (2013).

The local volatility of the price process, measured on a time interval τ small enough to neglect the evolution of $\varphi(t)$, is given by

$$\sigma^2(t) := \frac{1}{\tau}\mathbb{V}\left[\int_t^{t+\tau} \mathrm{d}m(t')\right] = \frac{1}{\tau}\mathbb{E}[N(t+\tau) - N(t)] = \vartheta^2\varphi(t),$$

where we have used the fact that $\mathbb{E}[\varepsilon] = 0$ and

$$\mathbb{E}[\varepsilon(t)\varepsilon(t')] = 0 \text{ for all } t' \neq t.$$

In other words, as in Section 2.2.4, volatility σ and market activity φ are one and the same thing in this simple framework (see Equation (2.18)).

For a Hawkes process, the intensity $\varphi(t)$ is a random variable with some distribution and some temporal correlations described by the function $c(t)$. This is an enticing model of prices, since its volatility is random and clustered in time, which matches empirical observations. In fact, choosing a power-law kernel Φ allows one to generate the **long-range correlations** that we discussed in Chapter 2.

Is this power-law Hawkes process for prices a simple way to reconcile these empirical observations in a parsimonious theoretical framework? As it turns out, there are both pros and cons to this idea. The fact that $\varphi(t)$ has long-range memory that can be described by a near-critical power-law kernel $\Phi(t)$ (see Section 9.2.5) is a clear plus. However, the distribution of φ values generated by a Hawkes process has relatively thin tails, which decay exponentially for large values of φ.[11]

This **exponential tail** is transmitted to the distribution of price changes, $m(t+\tau) - m(t)$, which therefore does not exhibit the power-law tails that are universally observed in real markets. In other words, Hawkes processes do not produce sufficiently large fluctuations in the activity rate to reproduce the large price moves induced by large activity bursts.

There are two (possibly complementary) paths to solving these issues with the model. The first path is to argue that Equation (9.31) is too restrictive, and to allow the amplitude of individual jumps to have fat tails (rather than simply taking the two values $\pm\vartheta$). This approach is certainly plausible for small-tick assets, for which both the bid–ask spread and the gap between the best and second-best occupied price levels in the LOB typically undergo large fluctuations. However, it is less satisfactory for large-tick assets, for which prices seldom change by more than one tick at a time but for which the distribution of returns aggregated over a longer time interval τ (say five minutes) still exhibits power-law tails. Moreover, this approach is not very satisfying, because it requires us to simply input a power-law

[11] In the limit $g = 1 - \epsilon$, with $\epsilon \to 0$, the full distribution of φ can be obtained as

$$P_{\mathrm{st.}}(\varphi) \propto \varphi^{\frac{2\varphi_0}{\omega}} e^{-2\epsilon\varphi/\omega},$$

which shows explicitly the exponential behaviour of the tail. More generally, the tail decays as $e^{-z_c\varphi}$, where z_c is the solution of $g\omega z_c = \ln(1 + \omega z_c)$.

distribution for individual price changes without understanding where it comes from.

The second path is to extend the model and allow the level of market activity to be influenced by the amplitude of price changes themselves, and not only by their frequency. This framework produces a much richer structure, in which power-law tails are indeed generated. We will discuss this approach – of **generalised Hawkes processes** – in the next section.

9.5 Generalised Hawkes Processes

We now consider some generalisations to the standard Hawkes framework that we have developed so far in this chapter, and also present some examples to illustrate how these generalisations can help with practical applications.

Inhibition

The condition that a Hawkes process feedback kernel $\Phi(t)$ is positive for all lags is both necessary and sufficient to ensure that the rate $\varphi(t)$ always remains positive for any realisation of the underlying Poisson process. A positive kernel always leads to self-excitation and clustering.

In some circumstances, however, it is desirable to describe **inhibitory effects**, such that previous events produce negative feedback on $\varphi(t)$. In the presence of such inhibitory effects, one way to ensure that $\varphi(t)$ remains positive is to write

$$\varphi(t) = F\left[\varphi_0 + \int_0^t \mathrm{d}N(u)\Phi(t-u)\right], \tag{9.32}$$

where $F(x)$ is a non-negative function such as $F(x) = \Theta(x)$ or $F(x) = \ln(1 + e^x)$. Provided that the slope of the function $F(x)$ is always less than or equal to 1, the same stability condition ($g < 1$) can be established in these cases.[12] Such models can be simulated and calibrated numerically, but are much harder to characterise analytically than Hawkes models that do not include such inhibitory effects.

Exogenous Events

Sometimes, it is known in advance that certain events occurring at known times (such as the release of scheduled macroeconomic news announcements) will lead to an increase in the activity rate. This effect is not a self-excitation, but is instead truly exogenous. A simple way to include these **exogenous** events in the Hawkes description is to write

$$\varphi(t) = \varphi_0 + \int_0^t \mathrm{d}N(u)\Phi(t-u) + \sum_{t_i<t} a_i \Phi_{\text{exo.}}(t-t_i), \tag{9.33}$$

where t_i is the time of the i^{th} such event, which has strength a_i, and $\Phi_{\text{exo.}}$ is another kernel describing how the impact of these events dies out with time.[13]

[12] See: Brémaud, P., & Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability*, 24, 1563–1588.

[13] See: Rambaldi, M., Pennesi, P., & Lillo, F. (2015). Modelling foreign exchange activity around macroeconomic news: Hawkes-process approach. *Physical Review E*, 91(1), 012819.

Multi-Event Hawkes Processes

The temporal evolution of an LOB is a multi-event problem, in which the different types of events can influence the future occurrences of both themselves and each other. For example, market order arrivals can trigger other market order arrivals, limit order arrivals, and cancellations (and vice-versa). The events that occur in other markets could also influence the activity in a given LOB.

The Hawkes framework can easily be extended to a *multivariate framework* with K event types, each with a time-dependent rate $\varphi_i(t)$, $i = 1,\ldots,K$. The multivariate form of Equation (9.10) is simply

$$\varphi_i(t) = \varphi_{0i} + \int_0^t \sum_{j=1}^K \mathrm{d}N_j(u)\Phi_{i,j}(t-u), \tag{9.34}$$

where $\Phi_{i,j}(t) \geq 0$ is a non-negative kernel function that describes how the past occurrence of an event of type j at time $t-u$ increases the intensity of the events of type i at time t.

The multivariate Hawkes framework is very versatile, because both the cross-event excitations and the temporal structures of the feedback can be chosen to reflect the activity observable in empirical data. The kernel $\Phi_{i,j}(t)$ need not be symmetric, so that the influence of type-i events on type-j activity need not be equal to the influence of type-j events on type-i activity.

In the case where the exogenous event rates are constant, the first- and second-order statistics can again be calculated exactly:

- The mean intensity of type-i events is given by a vector generalisation of the 1 component case, Equation (9.13):

  $$\bar{\varphi}_i = \sum_{j=1}^K [(\mathbf{1}-\mathbf{g})^{-1}]_{i,j}\varphi_{0,j},$$

  where $\mathbf{g}$ is the matrix of kernel norms

  $$g_{i,j} := \int_0^\infty \mathrm{d}u\, \Phi_{i,j}(u).$$

  The process is stable provided the spectral radius of the matrix $\mathbf{g}$ is strictly less than 1.
- For $t > u$, the conditional mean intensity is given by

  $$\mathbb{E}[\mathrm{d}N_i(t)|\mathrm{d}N_j(u) = 1] = \bar{\varphi}_i(1 + c_{i,j}(t-u))\mathrm{d}t,$$

  where $\mathbf{c}(u)$ is the solution of a matrix Wiener–Hopf equation

  $$\bar{\varphi}_i c_{i,j}(u) = \Phi_{i,j}(u) + \sum_{k=1}^K \int_0^\infty \mathrm{d}v\, \Phi_{ik}(u-v)\bar{\varphi}_k c_{kj}(v). \tag{9.35}$$

  Again, the lag-dependent matrix $\mathbf{c}$ is directly related to the covariance of the intensity as:

  $$\mathrm{Cov}\left[\frac{dN_i(t)}{dt}, \frac{dN_j(u)}{du}\right] = \bar{\varphi}_i\bar{\varphi}_j c_{i,j}(t-u) + \bar{\varphi}_i\delta_{i,j}\delta(t-u).$$

Hawkes Processes with Price Feedback

In a Hawkes process, past events feed back into future activity. Taking a step back from formalism and thinking about financial markets, it seems reasonable to conclude that while all types of market activity might influence future trading, price moves are particularly salient events that affect traders' future behaviour and decisions. For example, local trends (such as sequences of consecutive price changes in the same direction) induce reactions from the market, either because traders believe that these trends reveal some hidden information, or because if prices reach a certain level, they can trigger other execution decisions (such as stop-losses).

From this point of view, it is natural to extend the Hawkes formalism by writing:[14]

$$\varphi(t) = \varphi_0 + \int_0^t \mathrm{d}N(u)\Phi(t-u) + \left[\int_0^t \mathrm{d}m(u)\Psi(t-u)\right]^2, \qquad (9.36)$$

and

$$\mathrm{d}m(t) = \varepsilon(t)\vartheta \mathrm{d}N(t), \qquad (9.37)$$

where Equation (9.36) describes the intensity, Equation (9.37) describes the mid-price change, and Ψ is a kernel that measures the influence of past returns. In the simplest case, where Ψ is an exponential kernel, the trend is an exponential moving average of the recent price changes over some time horizon. The last term is the square of the trend, in which case only the amplitude (but not the sign) of the recent price change is relevant. This means that a large price change in a single direction produces a much greater increase in activity than does a sequence of many prices changes in alternating directions.

In this extended model with **price feedback**, one can derive closed-form expressions for the first-, second- and third-order statistics. For example, the mean activity rate $\bar{\varphi}$ is given by

$$\bar{\varphi} = \frac{\varphi_0}{1 - g_\Phi - g_{\Psi^2}},$$

where g_Φ is the norm of Φ and g_{Ψ^2} is the norm of the square of Ψ. The model thus becomes unstable when $g_\Phi + g_{\Psi^2} > 1$.

The specific form of Equation (9.36) is in fact motivated by empirical data, and the resulting model can be seen as a natural continuous-time definition of the usual GARCH model. Remarkably, when both Φ and Ψ are exponential kernels, some analytical progress is possible, and one can show that the distribution of φ acquires power-law tails. The value of the tail exponent depends on the norms g_Φ and g_{Ψ^2}, and can be set in the range of empirical values, even with a small value of the norm of g_{Ψ^2}. In other words, while the pure Hawkes model has exponentially decaying tails (and is therefore a poor starting point for explaining the distribution of returns), incorporating a small trend feedback allows one to generate much more realistic price trajectories with fat tails.

9.6 Conclusion and Open Issues

In this chapter, we have introduced the popular family of Hawkes models, which help to address the difficult problem of understanding the highly clustered and intermittent evolution of financial markets in calendar time. In a Hawkes process, the event arrival intensity at a given time depends not only on an exogenous component, but also on an endogenous or "self-excitation" term.

[14] See: Blanc, P., Donier, J., & Bouchaud, J. P. (2016). Quadratic Hawkes processes for financial prices. *Quantitative Finance*, 17, 1–18.

Hawkes processes are extremely versatile tools for modelling univariate and multivariate point processes in calendar time. Like Gaussian processes, Hawkes processes are entirely determined by their first- and second-order statistics. It is therefore relatively easy to calibrate a Hawkes process on data. Another benefit of using Hawkes processes is the possibility to interpret their fitted parameter values. This allows one to quantify interesting aspects such as the temporal structure of the memory kernel and the total feedback strength g.

When Hawkes models are calibrated on financial data, one finds empirically that the memory kernel has a power-law structure, which shows that multiple different time scales are needed to account for how markets respond to past activity. This is in line with many other observations, such as the well-known long memory structure that we discuss in Chapter 10. Note however that calibrating a Hawkes process to data does not prove the existence of a genuine causal self-excitation mechanism: as often, two-point correlations (used to calibrate the model) do not imply causality.

Perhaps surprisingly, empirical calibration systematically suggests that the feedback strength g is close to the critical value $g_c = 1$, beyond which the feedback becomes so strong that the activity intensity diverges and the Hawkes process ceases to be well defined. There are two possible ways to interpret this observation: one is that markets operate in a regime that is very close to being unstable; the other is that the Hawkes framework is too restrictive and fails to capture the complexity of real financial markets. In fact, as we emphasised in Section 9.2.6, the only way that a Hawkes process can reproduce long memory in the activity is for it to be critical, which is somewhat suspicious. More general models, such as the Hawkes model with price feedback, are not so constrained. This hints of a warning that is more general in scope: even when they are enticing and provide extremely good fits to the data, models can lead to erroneous conclusions because they fail to capture the underlying reality.

Take-Home Messages

(i) Market activity is not time-homogeneous, but rather exhibits both intra-day patterns and endogenous intermittency.

(ii) Real-time market activity can be modelled using point processes, in which events occur in continuous time according to a given rate (intensity). The simplest PP is a homogeneous Poisson process, for which inter-arrival times are independent exponential random variables with a constant rate parameter.

(iii) Linear Hawkes processes are simple auto-regressive models in which the intensity has both an exogenous component and an endogenous component. The exogenous component can vary with time to account for known trends or seasonalities. The endogenous component is measured by a kernel that describes how past events influence the present intensity.

(iv) The larger the norm of the kernel, the more "self-exciting" the process. The branching ratio describes the expected number of future events that will originate from a given event in a Hawkes process.

(v) When calibrated on financial data, the branching ratio is found to be close to 1, which suggests that a large fraction of market activity is endogenous.

(vi) Linear Hawkes processes fail to reproduce the strong volatility clustering and fat-tailed returns distributions that occur in real data. The Hawkes model can be made to produce more realistic returns by including additional effects, such as including price returns in the feedback mechanism.

9.7 Further Reading

General

Daley, D. J., & Vere-Jones, D. (2003). An introduction to the theory of point processes (Vols. I–II). *Probability and its applications*. Springer, second edition.

Bauwens, L., & Hautsch, N. (2009). Modelling financial high frequency data using point processes. In Andersen, T. G., Davis, R. A., Kreiss, & J.-P., Mikosch, Th. V. (Eds.), *Handbook of financial time series* (pp. 953–979). Springer.

Bacry, E., Mastromatteo, I., & Muzy, J. F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01), 1550005.

Intermittent Dynamics (Physics)

Alessandro, B., Beatrice, C., Bertotti, G., & Montorsi, A. (1990). Domain-wall dynamics and Barkhausen effect in metallic ferromagnetic materials. I. Theory. *Journal of Applied Physics*, 68(6), 2901–2907.

Frisch, U. (1997). *Turbulence: The Kolmogorov legacy*. Cambridge University Press.

Fisher, D. S. (1998). Collective transport in random media: From superconductors to earthquakes. *Physics Reports*, 301(1), 113–150.

Sethna, J. P., Dahmen, K. A., & Myers, C. R. (2001). Crackling noise. *Nature*, 410(6825), 242–250.

Mathematical Properties of Hawkes Processes

Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society*. Series B (Methodological), 33, 438–443.

Hawkes, A. G., & Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(03), 493–503.

Brémaud, P., & Massoulié, L. (2001). Hawkes branching point processes without ancestors. *Journal of Applied Probability*, 38(01), 122–135.

Bacry, E., Delattre, S., Hoffmann, M., & Muzy, J. F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7), 2475–2499.

Jaisson, T., & Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, 25(2), 600–631.

Bacry, E., & Muzy, J. F. (2016). First-and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory*, 62(4), 2184–2202.

Estimation of Hawkes Processes

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

Bacry, E., Dayri, K., & Muzy, J. F. (2012). Non-parametric kernel estimation for symmetric Hawkes processes: Application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems*, 85(5), 1–12.

Dassios, A., & Zhao, H. (2013). Exact simulation of Hawkes process with exponentially decaying intensity. *Electronic Communications in Probability*, 18(62), 1–13.

Hardiman, S. J., & Bouchaud, J. P. (2014). Branching-ratio approximation for the self-exciting Hawkes process. *Physical Review E*, 90(6), 062807.

Lallouache, M., & Challet, D. (2016). The limits of statistical significance of Hawkes processes fitted to financial data. *Quantitative Finance*, 16(1), 1–11.

Financial Applications of Hawkes Processes

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

Hardiman, S., Bercot, N., & Bouchaud, J. P. (2013). Critical reflexivity in financial markets: A Hawkes process analysis. *The European Physical Journal B*, 86, 442–447.

Bacry, E., & Muzy, J. F. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7), 1147–1166.

Da Fonseca, J., & Zaatour, R. (2014). Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 34(6), 548–579.

Achab, M., Bacry, E., Muzy, J. F., & Rambaldi, M. (2017). Analysis of order book flows using a nonparametric estimation of the branching ratio matrix. arXiv:1706.03411.

Extensions of Hawkes Processes

Brémaud, P., & Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24, 1563–1588.

Embrechts, P., Liniger, T., & Lin, L. (2011). Multivariate Hawkes processes: An application to financial data. *Journal of Applied Probability*, 48(A), 367–378.

Bormetti, G., Calcagnile, L. M., Treccani, M., Corsi, F., Marmi, S., & Lillo, F. (2015). Modelling systemic price cojumps with Hawkes factor models. *Quantitative Finance*, 15(7), 1137–1156.

Rambaldi, M., Pennesi, P., & Lillo, F. (2015). Modelling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E*, 91(1), 012819.

Blanc, P., Donier, J., & Bouchaud, J. P. (2016). Quadratic Hawkes processes for financial prices. *Quantitative Finance*, 17, 1–18.

Rambaldi, M., Bacry, E., & Lillo, F. (2016). The role of volume in order book dynamics: A multivariate Hawkes process analysis. *Quantitative Finance*, 17, 1–22.