



Testing for Self-excitation in Financial Events: A Bayesian Approach

Ali Caner Türkmen^(✉) and Ali Taylan Cemgil

Department of Computer Engineering, Boğaziçi University, Bebek,
34342 Istanbul, Turkey
{caner.turkmen,taylan.cemgil}@boun.edu.tr

Abstract. Self-exciting temporal point processes are used to model a variety of financial event data including order flows, trades, and news. In this work, we take a Bayesian approach to inference and model comparison in self-exciting processes. We discuss strategies to compute marginal likelihood estimates for the univariate Hawkes process, and describe a Bayesian model comparison scheme. We demonstrate on currency, cryptocurrency and equity limit order book data that the test captures excitatory dynamics.

1 Introduction

Many real-world data mining applications, including those in finance, entail modeling event occurrences in a continuous time setting. Examples of such data abound in finance; including order flows [3], trades [1], news [12], price jumps, volatility spikes, etc. Temporal point processes, statistical models of points scattered along the real line, are often the primary models used to address these data sets.

The Poisson process (PP) is one such statistical model that assumes independence among occurrences. Points are assumed to occur without any interaction, sometimes described as *completely randomly* [6]. PPs have been used in finance for modeling discrete event systems, e.g. limit orders [3]. While PPs lead to convenient mathematics for computing many quantities of interest analytically, they fail our simple intuition that financial events are seldom independent of one another, i.e. that they *excite* each other.

Self-exciting point processes, specifically Hawkes processes (HP) [7], are recently growing more common in quantitative finance [2] as well as machine learning literatures [8,9]. First explored in the backdrop of seismology, HPs assume causal, linear non-negative excitation behavior among occurrences. This is why they have been considered especially suited to modeling financial discrete events.

Typically, HPs are applied towards prediction tasks. Maximum likelihood estimates of model parameters are fit to an observation, a collection of occurrence timestamps, that are assumed to arise from the process. Model validation or

selection is then performed through predictive likelihood, or some other cross-validation metric, used to determine how good the fit is on a held out sample. Here, instead, we present a method of model selection (or equivalently, hypothesis testing) for self-exciting point process models. We take a Bayesian approach, and describe approximate inference and marginal likelihood estimation schemes. We present preliminary experiments on high frequency currency, cryptocurrency and equity limit order book data. Among a family of Bayesian inference methods, we posit that Laplace approximation to model evidence is best suited to the problem at hand.

In Sect. 2 we first give a brief overview of self-exciting processes and Bayesian model selection before describing our inference scheme. In Sect. 3, we present a set of preliminary findings on currency price, equity order book, and cryptocurrency event sets, before concluding in Sect. 4.

2 Model

2.1 Hawkes Process

Let $\{N(t)\}_{t \in \mathbb{R}_+}$ denote a counting process, a jump process where jump sizes are $+1$ and $N(0) = 0$. Furthermore, we will use the overloaded notation $N(a, b]$ to refer to the number of jumps (or equivalently, *points*) in the interval $(a, b]$ – also a random variable. In correspondence to a temporal point process, we think of $N(t)$ as the number of points –event occurrences such as orders or transactions– until time t .

Homogeneous Poisson processes are characterized by complete independence and stationarity assumptions. We have that $N(a, b]$ and $N(c, d]$ are independent random variables given that $(a, b]$ and $(c, d]$ are disjoint intervals on the real line. Furthermore, by stationarity we have that $\langle N(a, b] \rangle = \langle N(a + \tau, b + \tau] \rangle$ for all τ , where we let $\langle \cdot \rangle$ denote the expectation operator. However, it is these two assumptions that limit a realistic modeling of sequences of events that might as well have influenced each other.

Working with general classes of point processes where point occurrences are interdependent is difficult – both theoretically and computationally [6]. One alternative that leads to both mathematical and computational convenience is a class of temporal point processes (or, equivalently, counting processes), determined by a *conditional intensity* function [6]. Concretely, let λ^* denote the conditional intensity function of a self-exciting point process¹, defined by

$$\lambda^*(t) \triangleq \lim_{\delta \downarrow 0} \delta^{-1} \langle N(t, t + \delta] | \mathcal{H}_t \rangle.$$

Here we use \mathcal{H}_t to denote the *history* of events up to time t ². Note that setting $\lambda^*(t) = \nu(t)$, a deterministic measurable function of t , would simply yield a (nonhomogeneous) Poisson process.

¹ We follow the notation λ^* of [6], where the superscript $*$ serves as a reminder that the intensity function is dependent on the history up to time t , \mathcal{H}_t .

² Formally, \mathcal{H}_t can be seen as the natural filtration, an increasing sequence of σ -algebras, with respect to which we define the conditional expectation operator.

HPs arise as one of the simplest examples of point processes defined through a conditional intensity [4, 6]. They model *linear self-excitation* behavior, where the instantaneous probability of an event occurrence is given by a linear combination of the effects of past events. A (univariate) HP is a point process determined by the conditional intensity function [6, 7].

$$\lambda^*(t) = \mu + \sum_{t_j < t} \varphi(t - t_j). \quad (1)$$

Here $\mu > 0$ is the constant background (exogenous) intensity function. $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is the *triggering kernel*, an often monotonically decreasing function that governs self-excitation.

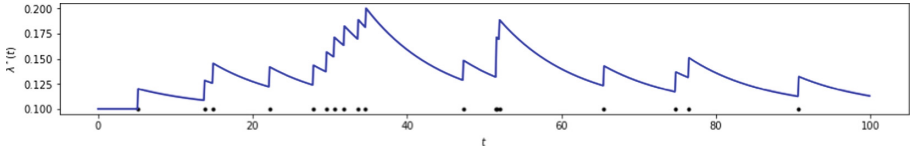


Fig. 1. Intensity function of a Hawkes process with exponential delay density

We will be concerned with the case $\varphi(x) = \alpha\theta \exp(-\theta x)$, where $\alpha \in [0, 1], \theta > 0$. Here since $\int \theta \exp(-\theta x) d\theta = 1$, we can interpret the triggering kernel in terms of its parameters. α governs the *infectivity* or the average number of new events that are *triggered* by an event. The remaining part is the exponential density for the length of the delay between events triggering each other. Note that $\alpha < 1$ is required for stationarity.

One can think of the intensity as a stochastic process itself, which is *excited* every time a jump occurs stochastically on the underlying process $N(t)$. That is, a jump in $N(t)$ leads to a jump of size α in λ^* . This effect then decays according to a schedule determined by the decay factor in φ , which in the case above, was taken as an *exponential* decay proportional to $\exp(-\theta \Delta t)$. We illustrate this effect in Fig. 1.

We refer the reader to the review by Bacry et al. [2] for further details on HP and their varied applications in quantitative finance.

Finally, let us note that for any conditional intensity point process the likelihood of finitely many points $\Pi = \{t_i\}_{i=1}^N$ where $0 < t_1 < \dots < t_N < T$ on a bounded interval $(0, T]$ is given by

$$p(\Pi|\lambda^*) = \exp\left(-\int_0^T \lambda^*(s) ds\right) \prod_{i=1}^N \lambda^*(t_i),$$

where the conditional intensity function $\lambda^*(x)$ uniquely determines the process. For Poisson processes, granted that the *compensator* $-\int_0^T \lambda(s) ds$ can be computed, the evaluation of the likelihood is trivial. This is not the case in general,

however. Note that the computation of the likelihood for a general HP defined as in (1) would take time $O(N^2)$, as each intensity evaluation takes time linear in the number of events. This crucial aspect prohibits the use of likelihood-based inference, including many Bayesian methods, in general. In the exponential kernel HP case, however, both the log likelihood and its gradient can be computed in linear time owing to the *memoryless* property. In the sequel, we constrain our attention to HP parameterized as such.

2.2 Bayesian Model Comparison

As mentioned previously, point processes are used mainly as models of discrete events occurring asynchronously in continuous time. Compared to discrete-time models that are often used in econometrics or time series forecasting, the methods of comparing and selecting models are less obvious.

Although HPs have been explored widely in finance, existing works often use cross-validation – basing model comparison on predictive likelihood, or other domain-driven measures of error on held out data. On the other hand, there is earlier work on frequentist hypothesis testing of HP vs PP [5]. In this paper, we present work in progress regarding a Bayesian approach – bringing the advantages (and potential pitfalls) of encoding prior assumptions on model parameters and deriving intuitive tests of model validity.

In Bayesian model comparison, one judges models through *marginal (integrated) likelihoods*, using the same calculus of probability that one judges parameter configurations of a fixed model. Let $p(\Pi|\Theta)$ denote the data likelihood, and $p(\Theta)$ a prior distribution under a certain model. Our aim is to compute the marginal likelihood

$$p(\Pi) = \int p(\Pi|\Theta)p(\Theta)d\Theta,$$

where we let Θ denote the vector of all model parameters. Intuitively, this quantity can be read as $\langle p(\Pi|\Theta) \rangle_{p(\Theta)}$, i.e. the expected likelihood that a given model will assign to data Π , as parameters are drawn from the prior $p(\Theta)$. Note that this quantity comes with “Occam’s razor” included, i.e. high-dimensional models with diffuse priors are automatically penalized. One can then use the marginal likelihoods of two different models to compare them.

Let p_1, p_0 denote marginal likelihoods under two different models. The ratio

$$BF = \frac{p_1(\Pi)}{p_0(\Pi)} \quad (2)$$

is known as the *Bayes factor*. Bayesian hypothesis tests are performed by calculating the marginal likelihood under the null (p_0), as well as the alternative (p_1) hypotheses, and computing BF . $BF > 10$ is taken as strong evidence that the first model (p_1) better explains the observations. Similarly, many models (or prior configurations) can be compared on the same footing.

2.3 Proposed Method

Here we propose a simple hypothesis test for “self-excitation” behavior in financial events. We calculate the Bayes factor (2) by taking a homogeneous PP as the null hypothesis (p_0), and an exponential-decay HP as given in (1) as the alternative (p_1). In doing so, we explore methods of marginal likelihood estimation for HP, which also paves the way to comparing HP models.

We equip both models (p_0, p_1) with appropriate prior distributions. In the former, we choose a Gamma distribution for the constant intensity parameter. The Gamma distribution is conjugate to the PP likelihood, making marginal likelihood computation analytically tractable. For HP, parameters μ, α, θ are given Gamma, Beta and Gamma priors respectively.

Marginal likelihood for HPs is intractable under any choice of prior, and we must resort to an approximation. Yet, this approximation is still made difficult by computational challenges related to the likelihood, outlined above. For example, one sampling-based alternative for marginal likelihood estimation, annealed importance sampling [11], requires a large number of likelihood computations before a single weighted sample can be drawn. This prohibits a realistic application of this method for HPs with large observed samples.

However, especially in the high-frequency context, we can invoke another approximation method. Financial continuous time data sets, unlike earthquakes, are characterized by large sample sizes. We find that this leads to peaked, unimodal posteriors, with which we can turn to Laplace approximation to the marginal likelihood [10].

We approximate the posterior with a multivariate Gaussian distribution centered around the posterior mode, $\Theta^* = \arg \max p(\Theta|\Pi)$. Given the posterior potential $\varphi(\Theta) = p(\Theta|\Pi)p(\Theta)$, we approximate $p(\Pi) = \int d\Theta \varphi(\Theta)$ via

$$\ln p(\Pi) \approx \ln \varphi(\Theta^*) + \frac{3}{2} \ln 2\pi - \frac{1}{2} \ln |H|,$$

where $H = \nabla^2 - \varphi(\Theta^*)$ is the Hessian of $-\varphi$ evaluated at the mode.

This method reduces marginal likelihood estimation to a series of simple steps. First, maximum a posteriori (MAP) estimates of HP are obtained. This can be achieved via expectation maximization, as well as gradient-based methods in the simple case of univariate HP. The Hessian H can be approximated numerically or computed exactly. Software for estimating marginal likelihood, as well as other tasks such as posterior inference under univariate Bayesian HP, is made available online³.

3 Experiments

Our experiments cover a range of financial event sets. FX are high frequency (millisecond range) tick events in an interbank currency exchange, previously

³ <http://www.github.com/canerturkmen/hawkeslib>, and on the Python Package Index (PyPI) as `hawkeslib`.



Fig. 2. Data samples from the three data sets. x-axis denotes time of occurrence, y-axis is random noise for better visibility

Table 1. Results of experiments on financial data sets. N denotes the number of occurrences in the data set. BF is the computed Bayes factor. Bayesian credible intervals at 95% are given for α, θ

Data Set	Asset	N	BF	α		θ	
				Lower	Upper	Lower	Upper
FX	GBPUSD	1000	$>10^8$	0.62	0.75	5.46	8.06
	EURUSD	1000	$>10^8$	0.53	0.64	6.26	8.37
	USDJPY	1000	$>10^8$	0.61	0.73	5.15	7.12
Crypto	USDT-ETH	5684	0.11	0.15	0.38	0	0.02
	BTC-XRP	5710	0.04	0.08	0.45	0	0.02
	BTC-ETH	5499	3.69	0.12	0.34	0	0.02
LOB	GARAN	1000	$>10^8$	0.88	0.99	8.8	18

investigated using HP [13]. We model three large-volume currency pairs selected at random. **Crypto** are price increase events on three large-cap cryptocurrencies on the cryptocurrency exchange Bittrex sampled at five minute (low) frequency. Finally **LOB** are limit order arrivals in a large-cap bank stock in the Turkish equity exchange, Borsa Istanbul, sampled at very high frequency (nanosecond range). Samples of each data set are given in Fig. 2. In **FX** and **LOB**, we limit event sets to 1000 events, roughly to 10 min of trading. Observe that in both data sets, the data cluster around certain points in time. This effect is less pronounced in **Crypto**.

We report the results of our tests, where we calculate the Bayes factor as described in Sect. 2.3. We further present 95% Bayesian credible intervals for the triggering kernel parameters, where we use simple random walk Metropolis (RWM) [10] algorithm to draw from the posterior.

We present the results in Table 1. The test accurately captures that low-frequency price jumps do not present sufficient evidence in favor of self-excitation. In **FX** and **LOB**, however, we find overwhelming evidence that HP outperforms PP. Note however that, if one were to register only large return

jumps as events, HPs could still fit the data at lower frequencies. This is not surprising, its analogue in the discrete-time setting would be known as *volatility clustering*.

There are, however, two issues we must address. First, Bayesian analysis is well known to be sensitive to choice of priors. In our analyses, we find that large data sets easily mitigate this effect. In Fig. 2, we change the scale hyperparameter of the prior for θ , the delay distribution. We find that, except for unrealistic choices of priors which set the average delay to less than 0.01 ms, the conclusion is largely unaffected. Varying other hyperparameters lead to similar conclusions.

Finally, let us note that this paper and many others in the field assume constant background intensity μ . The test in this paper also assumes a homogeneous PP as the null hypothesis. However, the *exogenous* process that governs financial events is often not stationary. For example, financial events follow intra-day, weekly and yearly cycles. Our test, and many other investigations in HP, are prone to capturing this effect and explaining it away using the *endogenous* component of HP. We test this effect using a toy data set drawn from a non-homogeneous PP with intensity $\lambda(t) \propto \exp(\sin t)$ (see, e.g. Figure 4). On this data, our test easily passes (rejects PP), although the nonstationarity is purely exogenous. In our experiments, we mitigate the potential effect of periodicity by sampling short time intervals (Fig. 3).

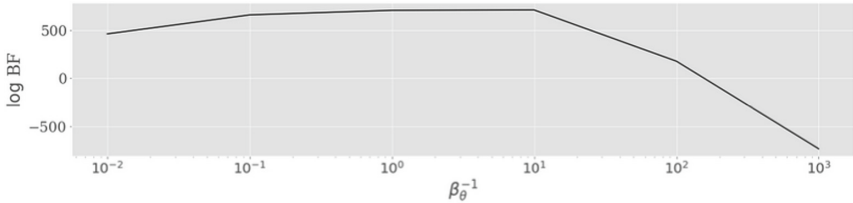


Fig. 3. Logarithm of the Bayes factor as the scale hyperparameter of $p(\theta)$ is changed; for EURUSD in the **FX** data set.

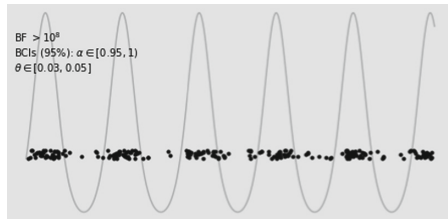


Fig. 4. A draw from a nonhomogeneous Poisson process with periodic intensity

4 Conclusion

We combined techniques from Bayesian machine learning and evolutionary point processes for modeling high-frequency financial data. We cast HP to a Bayesian setting, and discussed the computation of a Bayesian model comparison scheme for testing “self-excitation” behavior in financial events as well as posterior inference. Early experiments confirm basic intuition regarding high-frequency financial events.

Our method can be used to capture self-excitation effects in financial discrete event data, much in the same way conditional heteroskedasticity models capture volatility clustering. However, the test assumes that background intensities are stationary, and can lead to pitfalls in financial analysis. This issue constitutes the next step to this study.

Acknowledgement. We gratefully acknowledge the support of Scientific and Technological Research Council of Turkey (TUBITAK), under research grant 116E580.

References

1. Bacry, E., Dayri, K., Muzy, J.F.: Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *Eur. Phys. J. B* **85**(5), 157 (2012)
2. Bacry, E., Mastromatteo, I., Muzy, J.F.: Hawkes processes in finance. *Mark. Microstruct. Liq.* **1**(01), 1550005 (2015)
3. Cont, R.: Statistical modeling of high-frequency financial data. *IEEE Sig. Process. Mag.* **28**(5), 16–25 (2011)
4. Cox, D.R., Isham, V.: *Point Processes*, vol. 12. CRC Press, Boca Raton (1980)
5. Dachian, S., Kutoyants, Y.A.: Hypotheses testing: poisson versus self-exciting. *Scand. J. Stat.* **33**(2), 391–408 (2006)
6. Daley, D.J., Vere-Jones, D.: *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. PIA. Springer, New York (2003). <https://doi.org/10.1007/b97277>
7. Hawkes, A.G.: Point spectra of some mutually exciting point processes. *J. R. Stat. Soc. Ser. B (Methodol.)* **33**, 438–443 (1971)
8. Linderman, S., Adams, R.: Discovering latent network structure in point process data. In: *International Conference on Machine Learning*, pp. 1413–1421 (2014)
9. Mei, H., Eisner, J.M.: The neural Hawkes process: a neurally self-modulating multivariate point process. In: *Advances in Neural Information Processing Systems*, pp. 6757–6767 (2017)
10. Murphy, K.P.: *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge (2012)
11. Neal, R.M.: Annealed importance sampling. *Stat. Comput.* **11**(2), 125–139 (2001)
12. Rambaldi, M., Pennesi, P., Lillo, F.: Modeling FX market activity around macroeconomic news: a Hawkes process approach, vol. 67, p. 210. arxiv preprint. arXiv preprint [arXiv:1405.6047](https://arxiv.org/abs/1405.6047) (2014)
13. Türkmen, A.C., Cemgil, A.T.: Modeling high-frequency price data with bounded-delay Hawkes processes. In: Corazza, M., Durbán, M., Grané, A., Perna, C., Sibillo, M. (eds.) *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 507–511. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-89824-7_90