Branching ratio approximation for the self-exciting Hawkes process

Stephen J. Hardiman and Jean-Philippe Bouchaud
Capital Fund Management,
23 rue de l'Université, 75007 Paris, France

(Dated: March 25, 2014)

We introduce a model independent approximation for the branching ratio of Hawkes self-exciting point processes. Our estimator requires knowing only the mean and variance of the event count in a sufficiently large time window, statistics that are readily obtained from empirical data. The method we propose greatly simplifies the process of Hawkes branching ratio estimation, proposed as a proxy for market endogeneity in recent publications and formerly estimated using numerical maximisation of likelihood. We employ this method to support recent theoretical and experimental results indicating that the best fitting Hawkes model to describe S&P futures price changes is in fact critical (now and in the recent past) in light of the long memory of financial market activity.

I. THE SELF-EXCITING HAWKES PROCESS

The Hawkes model is a simple and powerful framework for simulating or modelling the arrival of clustered events. In one dimension, the model is a point process $N(t)$ with an intensity $\lambda(t) = E\left[\frac{dN(t)}{dt}\right]$ given by a constant term μ and a 'self-exciting' term which is a function of the event history.

$$\lambda(t) = \mu + \int_{-\infty}^{t} \phi(t-s)dN(s) \qquad (1)$$

$\phi(\tau) \geq 0$ is an "influence kernel" which decides the weight to attribute to past events. The base intensity μ and the kernel shape $\phi(t)$ are parameters to be varied. A popular choice for the kernel is the exponential function $\phi(\tau) = \alpha e^{-\beta\tau}$ [1, 2] but in general the kernel to be used should depend on the application or the dynamics of the data to be modelled.

It is easy to show that with the assumption of stationarity, the average event rate of the process defined in Equation (1) is $\Lambda = \mu/(1-n) \geq \mu$. One can create a direct mapping between the Hawkes process and the well known branching process [3] in which exogenous "mother" events occur with an intensity μ and may give rise to x additional endogenous "daughter" events, where x is drawn from a Poisson distribution with mean $n = \int \phi(\tau)$. These in turn may themselves give birth to more "daughter" events.

The value n, which corresponds with the integral of the Hawkes kernel is the branching ratio. If $n > 1$, meaning that each event typically triggers at least one extra event, then the process is non-stationary and may explode in finite time. However, for $n < 1$, the process is well behaved and has proven useful in modeling the clustered arrival of events in a wide variety of applications from finance to geophysics.

The self-exciting Hawkes process has recently been applied as a means of measuring market endogeneity or 'reflexivity' in financial markets [1, 4]. In [1], authors consider mid-price changes in the E-mini S&P Futures contract between 1998 and 2010 and observe that the branching ratio of the best fitting exponential model has been increasing steadily over this period. They argue that this observation indicates that the market has become more reflexive in recent years with the rise of high frequency and algorithmic trading and is therefore more prone to market instability and so-called "flash crashes".

In [5], however, we have argued that due to the presence of long-range dependence in the event rate of mid-price changes (detectable in both 1998 and 2011) that provided one studies a sufficiently large window of observation the best fitting stationary Hawkes model must in fact be critical, i.e. have a branching ratio $n = 1$. This is backed up by theoretical arguments and empirical measurements on market data.

II. MAXIMUM LIKELIHOOD ESTIMATION

Given observed events (e.g. mid-price changes) at times $t_1, t_2, \ldots, t_n$ in an interval $[0, T]$ we can fit the Hawkes model by maximising the log-likelihood [2, 6] over the set of parameters θ.

$$\log L(t_1, t_2, \ldots, t_n|\theta) = -\int_0^T \lambda(t|\theta)dt + \int_0^T \log \lambda(t|\theta)dN(t) \qquad (2)$$

In the case of the exponential kernel $\theta = \{\mu, \alpha, \beta\}$. In practice, the model parameters $\{\hat{\mu}, \hat{\alpha}, \hat{\beta}\}$ which maximise this log-likelihood are obtained with numerical techniques due to the lack of a closed form solution. The branching ratio estimate is then $\hat{n} = \hat{\alpha}/\hat{\beta}$.

However, there are a number of pitfalls to using this procedure as a means of estimating the Hawkes kernel integral $n = \int \phi(\tau)$ [7]. Arguably the most important of which is that any estimate of the branching ratio made in this manner will be heavily dependent on the choice of kernel model (e.g. exponential, power-law, etc.) It may be that no parameter choice with the model we choose can satisfactorily describe the events observed and

the meaning of the branching ratio extracted is therefore questionable.

Care must also be taken when employing this method in the presence of imperfect event data as illustrated in [7]. In one of their figures, the authors present a (negative) log-likelihood surface which features two minima (one local, and one global). The global -log-likelihood minimum does little more than describe packet clustering inside the millisecond which arises from the manner in which events that arrive at the exchange at different times are bundled and recorded with the same timestamp. Subsequent randomisation of the timestamps inside a short time interval (in this case, one millisecond) creates a spurious high frequency correlation. The parameters associated with this minimum produce a very high event intensity that lasts for less than a millisecond after the arrival of an event. This explains why many seconds may pass but then two events can arrive almost instantaneously. The local -log-likelihood minimum, which is in fact a better fit to the 'true' lower frequency dynamics does a poor job of explaining this spurious clustering and is punished with a lower log-likelihood. Indeed, when the authors of [7] choose to fit this local minimum they corroborate results presented in [5].

We believe it is therefore essential to have additional checks (such as non-parametric methods [8]) at one's disposal to support results obtained from likelihood maximisation. To address the pitfalls in branching ratio estimation that arise from the model choice we propose a simple model-independent tool for branching ratio estimation, in the next section, which accurately reproduces previous results of [1] and also indicates the criticality of the relevant Hawkes process which describes the market.

III. A MEAN-VARIANCE ESTIMATOR FOR THE BRANCHING RATIO n

We begin with a general expression relating the Fourier transform of the kernel function to the Fourier transform of the infinitesimal auto-covariance function of the event rate (see [8, 9] for a derivation).

$$\hat{\nu}(\omega) = \frac{\Lambda}{\left| 1 - \hat{\phi}(\omega) \right|^2} \qquad (3)$$

Setting $\omega = 0$ we obtain a relation between the branching ratio, the average event rate and the integral of the auto-covariance (in the stationary case $n \leq 1$)

$$\int_{-\infty}^{\infty} \nu(t)dt = \frac{\Lambda}{\left| 1 - \int_{-\infty}^{\infty} \phi(t)dt \right|^2} \qquad (4)$$

$$= \frac{\Lambda}{(1-n)^2} \qquad (5)$$

Therefore, to deduce the branching ratio of the stationary Hawkes process, we need only build an estimator for Λ and $\int_{-\infty}^{\infty} \nu(t)dt$. Estimating Λ is trivial. To estimate $\int_{-\infty}^{\infty} \nu(t)dt$ we consider the variance of event rate in a window of length W.

$$\text{Var}[N_W] = \int_{-W}^{W} (W - |t|)\, \nu(t)dt \qquad (6)$$

We then note that if:

1. $W \gg R$ such that $(W - |t|)/W \approx 1$ for all $|t| < R$

2. $v(t) \approx 0$ for $|t| > R$

then

$$\int_{-\infty}^{\infty} \nu(t)dt \approx \frac{\text{Var}[N_W]}{W} \qquad (7)$$

and we find

$$n \approx 1 - \sqrt{\frac{\text{E}[N_W]}{\text{Var}[N_W]}} \qquad (8)$$

The estimator has a very intuitive interpretation. When the variance is equal to event rate the process is obeying Poisson statistics and $n = 0$. Any increase in the variance above the event rate indicates correlation and, within a Hawkes framework, a positive branching ratio.

IV. EMPIRICAL APPLICATIONS

A. Flash-crash revisited

To demonstrate the effectiveness of this simple estimator we have repeated the flash-crash day branching ratio analysis of Filimonov & Sornette [1]. We consider non-overlapping periods of 10 minutes in the hours of trading before and just after the flash-crash. For each 10-minute period, we calculate the sample mean and variance of the number of mid-price changes in the 60 windows of length $W = 10s$ contained. The resulting values are plugged into Equation 8 to obtain an approximation of the branching ratio for each period. The results are presented in Figure 1.

Note that this simple estimator is only a biased approximation, and for a general W, will underestimate the value of $\int_{-\infty}^{\infty} \nu(t)dt$ and therefore the branching ratio. Since we consider a window size $W = 10s$ we have systematically underestimated n in Figure 1 as measurements of $\nu(t)$ on this data have support outside the interval $[-10s, 10s]$ — there is still significant autocorrelation in the event rate at scales longer than $10s$.

However, we have identified that a window size W of the order of approximately 10 to 30 seconds produces estimates of the branching ratio on our data in line with those obtained by ML fitting to the exponential model
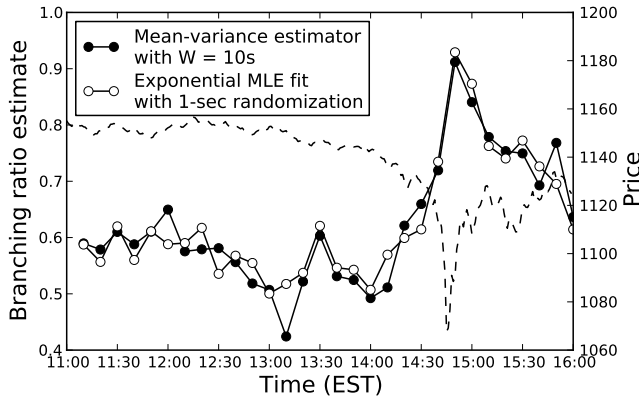
FIG. 1: Reproduction of the flash-crash branching ratio analysis of Filimonov & Sornette using the mean-variance estimator. Our results compare well with those obtained by maximising the likelihood of the exponential model (with 1-second randomisation). The dashed line is the E-mini S&P price. The points are placed at the end of the 10-minute period for which the branching ratio estimate is made.
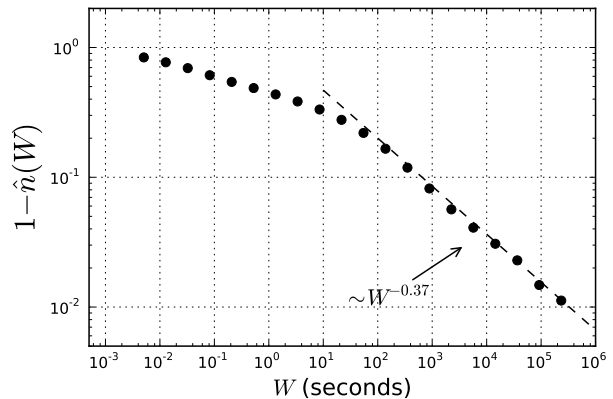

FIG. 2: $1 - \hat{n}$, the estimate of the branching ratio as a function of window size for E-mini S&P mid-price changes in 2010. The mean and variance of N_W are estimated on the full year. Note that we have 'stitched' together periods of regular trading hours (09:30 to 16:00 EST) and have de-trended the event count by the average U-shaped intra-day seasonality.

after intra-second randomisation (the method applied in [1].) Note that we do not fix β in our ML fit but let it settle naturally at the value which maximises the log-likelihood. We observe that this value $\hat{\beta}$ is very much dependent on the period of randomisation[1] of the timestamps. When we randomise timestamps inside each millisecond in we obtain $\hat{\beta}^{-1} \approx 10^{-2}$ for periods in 2010 but randomisation at larger intervals (e.g. the intra-second randomisation of Filimonov & Sornette) prevents $\hat{\beta}^{-1}$ from exceeding values of the order of magnitude of the scale of randomisation.

Note that since our results with $W = 10s$ correspond well with those obtained using the methods of Filimonov & Sornette [1], their procedure must *also* underestimate the branching ratio. To converge on the true n in expectation, we must take expression 8 in the limit of $W \to \infty$. We do just this in Figure 2 for mid-price changes of the E-mini S&P Futures contract in 2010. One notes that as the window size increases, the branching ratio converges to $n = 1$ in a non-trivial way. As reported in [5] for the structure of the kernel at short and long time-scales, two regimes are detectable with a transition around five minutes. The branching ratio asymptotically tends towards 1.0 with a scaling exponent $\epsilon = -0.37$ not too different from the value of 0.45 estimated in [5] for a 14-year period. Note that in taking the limit $W \to \infty$ we consider variation in the event rate at significant time-scales to be explained by the stationary Hawkes model.
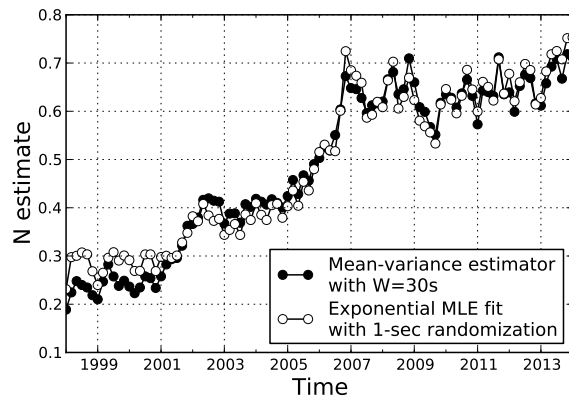

FIG. 3: Estimates of the branching ratio for 15-minute windows using the method of Filimonov & Sornette [1] and estimates using the mean-variance estimator with a window size of $W = 30s$. Note that our MLE results differ somewhat from the plot presented in [1]. We attribute this to differences in the data source or identification of mid-price changes.

B. Reflexivity : 1998 - 2013

Using the mean-variance estimator, we can also easily reproduce the result of [1] that claims to demonstrate that reflexivity has been increasing in the S&P futures market since 1998. To do this we set our only parameter $W = 30s$ and estimate the branching ratio in periods of 15-minutes. In Figure 3 we present the 2-month medians of these estimates beside the median of the branching ratio estimate obtained using the exponential maximum likelihood approach after intra-second randomisation.

Since we expect the minimum time-scale of correlation in the data to have decreased over the past decades

---

[1] To address the limited precision of the event data in [1], the authors randomise timestamps uniformly inside the second that they are reported.
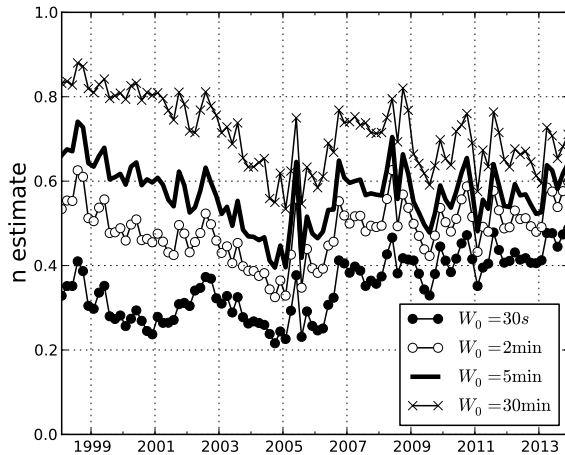
FIG. 4: Estimates of the branching ratio on 2-month periods using the mean-variance estimator with a window size that follows Moore's law: $W(t) = W_0 e^{-c(t-t_0)}$ with $c = -\log(1/2)/(18\,\text{months})$ and $t_0 = 1998$. We again stitch together periods of regular trading hours and de-trend the event count by the intra-day U-shape for each year. When W is appropriately rescaled, the branching ratio estimate is approximately constant through time, for all values of W_0.

(due to decreasing latency with advancing technology) we now re-perform the experiment with a window size W that follows Moore's law: the window size halves every 18 months (this describes well the increase in the high frequency activity of markets, see [5].) The results of this experiment are presented in Figure 4 and confirm that the kernel integral is approximately invariant over time provided that the measurement window W is appropriately rescaled to account for the changing speed of interactions in the market.

V. SUMMARY

We have introduced a simple estimator for the branching ratio of the Hawkes self-exciting point-process. The method is straight-forward to apply to empirical event data since it requires only a rudimentary calculation on the mean and variance of the event rate. The estimator does not suffer from the bias inherent to the contentious question of model selection in the likelihood maximisation approach and furthermore it avoids the need for complicated numerical minimisation techniques.

Despite its simplicity, our estimator can accurately reproduce results obtained for the branching ratio using this prior method. The estimator presented is indeed a biased estimator, but it is asymptotically unbiased in the limit of large W for which we observe that the branching ratio for empirical mid-price changes of the E-mini S&P futures contract approaches unity in line with previous theoretical and empirical results [5].

Furthermore we demonstrate that if the window size is allowed to scale with Moore's law and adapt to the changing speed of interaction in the market over the past fifteen years, then the branching ratio approximation recovered is constant further confirming prior observations of the invariance of the Hawkes kernel and branching ratio over time in [5].

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

[2] T. Ozaki, "Maximum likelihood estimation of hawkes self-exciting point process," *Annals of the Institute of Statistical Mathematics* **31** (1979) 145–155. Part B.

[3] T. E. Harris, *The Theory of Branching Processes.* Dover Publications, 2002.

[4] V. Filimonov, D. Bicchetti, N. Maystre, and D. Sornette, "Quantification of the high level of endogeneity and of structural regime shifts in commodity markets," *Journal of International Money and Finance* **42** (2014) 174–192.

[5] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud, "Critical reflexivity in financial markets: a hawkes process analysis," *The European Physical Journal B* **86** no. 10,

(2013) 1–9.

[6] I. Rubin, "Regular point processes and their detection," *IEEE Transactions on Information Theory* **IT-18** (September, 1972) 547–557.

[7] V. Filimonov and D. Sornette, "Apparent criticality and calibration issues in the hawkes self-excited point process model: application to high-frequency financial data," `arXiv:1308.6756 [q-fin.ST]`.

[8] E. Bacry, K. Dayri, and J.-F. Muzy, "Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data," *European Physics Journal B* **85** no. 5, (2012) 157.

[9] A. G. Hawkes, "Point spectra of some mutually exciting point processes," *Biometrika* **58** (1971) 83.