Mid-Price Movement Prediction using Hawkes Processes

Yuanda Chen Major Advisor: Dr. Alec Kercheval
Financial Mathematics, Florida State University

Contribution

We introduce a 4-dimensional Hawkes processes model to simulate the dynamics of Limit Order Book. Each dimension corresponds to the market events that increases/decrease the queue sizes at the best bid/ask. Comparing to the homogeneous Poisson process, the Hawkes process captures the well-observed clustering effect of each single kind of market event. At the meanwhile, it allows us to build-in the cross-excitation among different kinds of market events.

Monte Carlo simulation is used to estimate the probability of the next mid-price movement being upward. An algorithmic trading strategy is designed to evaluate the performance.

M-Dimensional Hawkes Process

An M-dimensional Hawkes process with exponential kernel has the conditional intensity:

$$\lambda_m(t|\mathcal{F}(t)) = \mu_m + \sum_{n=1}^{M} \alpha_{mn} \int_0^t e^{-\beta_{mn}(t-u)} dN_n(u).$$

Given the occurrence times $\{t_i^n\}_{n=1,2,\ldots,M}$,

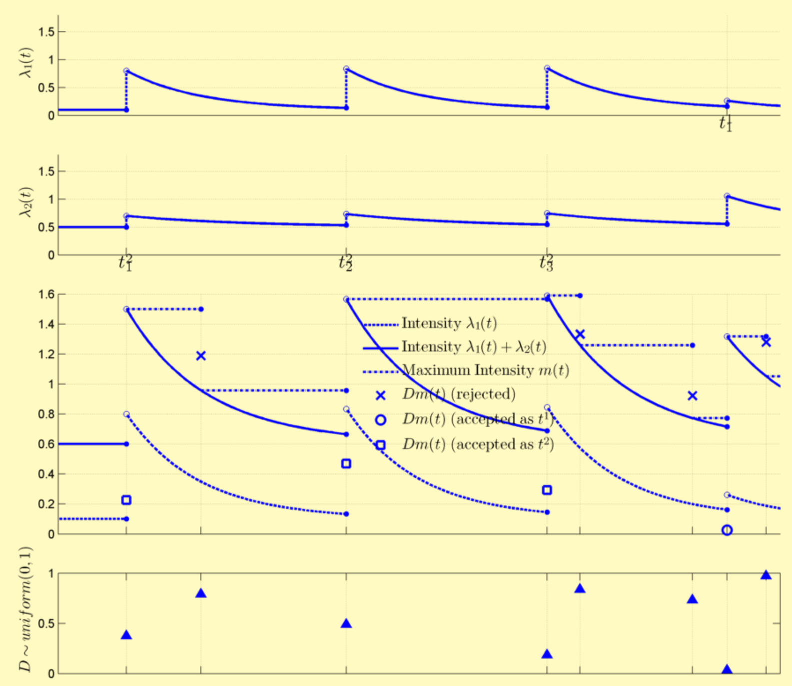
$$\lambda_m(t) = \mu_m + \sum_{n=1}^{M} \alpha_{mn} \sum_{\{k:t_k^n<t\}} e^{-\beta_{mn}(t-t_k^n)}.$$

for $m = 1, 2, \ldots, M$.

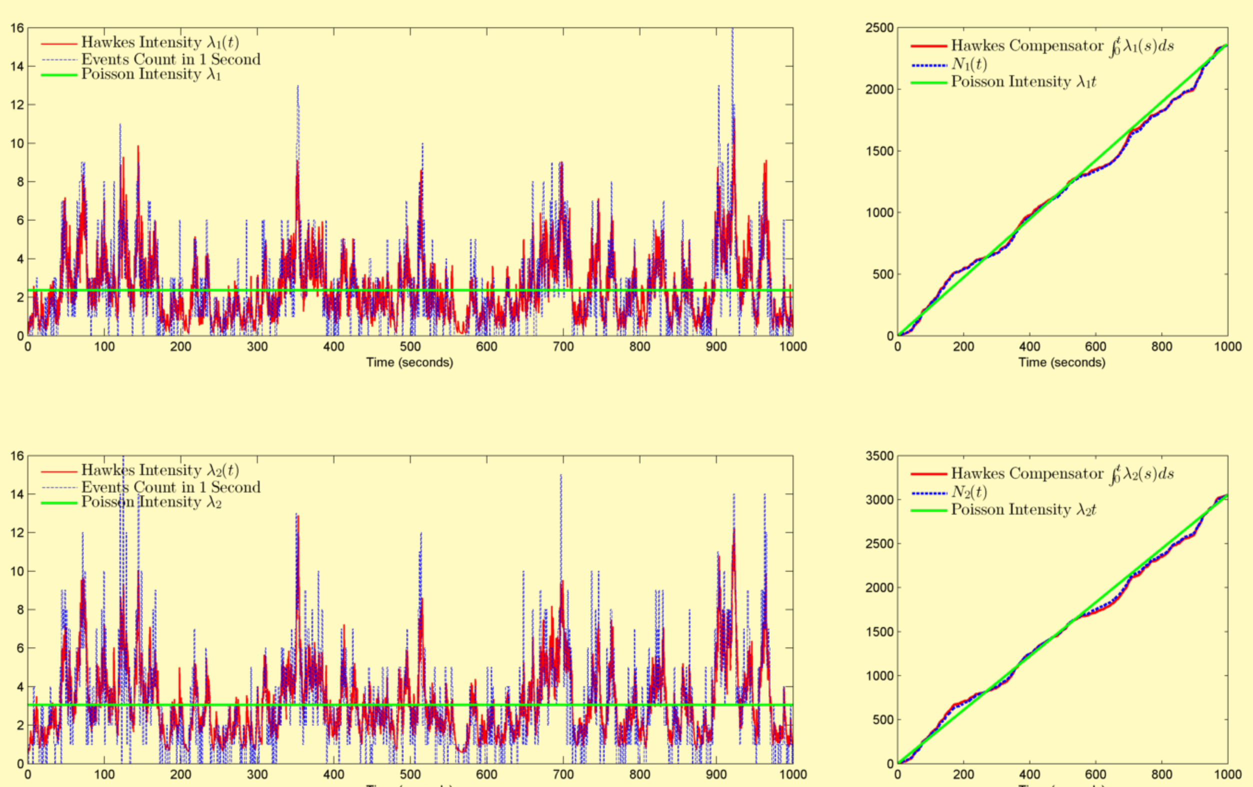
Simulating Hawkes Process

Algo: Thinning Algorithm

Input: U, A, B, T
Initialize $\{t^i\}_{i=1,2,\ldots,M} = \{\}$, $s = 0, m = \mathtt{sum}(U)$;
while $s < T$ **do**
  Draw $\Delta s \sim \mathtt{exponential}(m)$;
  Set $s = s + \Delta s$;
  Compute $\lambda = \mathtt{sum}(\lambda(s|\mathcal{F}_t))$;
  Draw $D \sim \mathtt{uniform}(0,1)$;
  if $D \leq \lambda/m$ **then**
    /* /w prob. proportional to intensity */
    Accept s to t^n ;
    /* column sum */
    Set $m = \lambda + \mathtt{sum}(A_{:,n})$;
  else
    Set $m = \lambda$;
  end
end
return $\{t^i\}_{i=1,2,\ldots,M}$



Hawkes v.s. Poisson



	μ_1	μ_2	α_{11}	α_{12}	α_{21}	α_{22}	β_{11}	β_{12}	β_{21}	β_{22}
True (blue)	0.1	0.5	0.1	0.7	0.5	0.2	1.2	1.0	0.8	0.6
Hawkes (red)	0.12	0.58	0.08	0.61	0.55	0.21	0.71	0.95	0.91	0.60
Poisson (green)	2.37	3.05	-	-	-	-	-	-	-	-

Calibration

The model parameters are calibrated by numerically optimizing the likelihood:

$$\log L\left(\{t_i^n\}_{n=1,2,\ldots,M}\middle|\theta\right) = \sum_{m=1}^{M} \log L_m\left(\{t_i^n\}_{n=1,2,\ldots,M}\middle|\theta\right)$$

$$= \sum_{m=1}^{M}\left\{-\mu_m T - \sum_{n=1}^{M}\frac{\alpha_{mn}}{\beta_{mn}}\sum_{\{k:t_k^n<T\}}\left[1-e^{-\beta_{mn}(T-t_k^n)}\right] + \sum_{\{k:t_k^m<T\}}\log\left[\mu_m + \sum_{n=1}^{M}\alpha_{mn}R_{mn}(k)\right]\right\}$$

with the recursion:

$$\begin{cases} R_{mn}(i) &= \sum_{\{k:t_k^n<t_i^m\}} e^{-\beta_{mn}(t_i^m-t_k^n)} = e^{-\beta_{mn}(t_i^m-t_{i-1}^m)}R_{mn}(i-1) + \sum_{\{t_{i-1}^m \leq t_k^n < t^m\}} e^{-\beta_{mn}(t_i^m-t_k^n)} \\ R_{mn}(0) &= 0 \end{cases}$$

Diffusion Approximation

Let $K = \{\frac{\alpha_{ij}}{\beta_{ij}}\}$, $\bar{\mu} = (I-K)^{-1}\mu T$, $\bar{\sigma} = (I-K)^{-1}\left[diag\left((I-K)^{-1}\mu\right)\right]^{1/2}\sqrt{T}$, then $N(tT) \xrightarrow{T\to\infty} \bar{\mu}t + \bar{\sigma}W_t$. The queue sizes at best bid/ask can be modeled as:

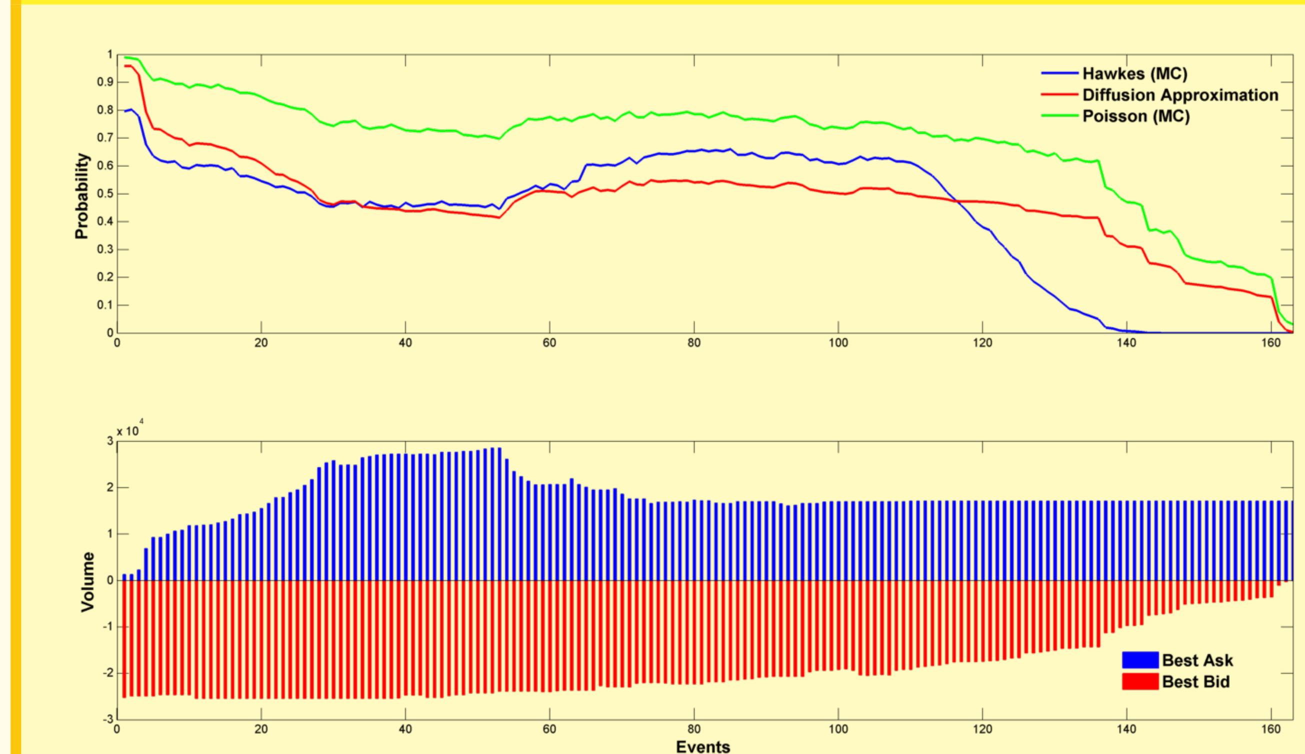
$$\begin{cases} Q^a(t) &= Q_0^a + S_L N^4(t) - S_M N^3(t) \\ Q^b(t) &= Q_0^b + S_L N^2(t) - S_M N^1(t) \end{cases}$$

Assuming no drift terms in Q^a and Q^b, we have a 2-dimensional correlated Brownian Motion.

$$\begin{cases} dQ^a(t) &= \sum_{i=1}^{4}(S_L\bar{\sigma}_{4i} - S_M\bar{\sigma}_{3i})dW_t^i \\ &= \sqrt{\sum(S_L\bar{\sigma}_{4i} - S_M\bar{\sigma}_{3i})^2}dW_t^1 \\ dQ^b(t) &= \sum_{i=1}^{4}(S_L\bar{\sigma}_{2i} - S_M\bar{\sigma}_{1i})dW_t^i \\ &= \sqrt{\sum(S_L\bar{\sigma}_{2i} - S_M\bar{\sigma}_{1i})^2}dW_t^2 \\ \rho &= \frac{\sum(S_L\bar{\sigma}_{4i}-S_M\bar{\sigma}_{3i})(S_L\bar{\sigma}_{2i}-S_M\bar{\sigma}_{1i})}{\sqrt{\sum(S_L\bar{\sigma}_{4i}-S_M\bar{\sigma}_{3i})^2}\sqrt{\sum(S_L\bar{\sigma}_{2i}-S_M\bar{\sigma}_{1i})^2}} \end{cases}$$

This allows fast approximation of the probability for an upward mid-price movement (i.e. $Q^a(t)$ becomes negative before $Q^b(t)$ does) from the analytical solution.

Results



This example shows an event-by-event calculation of the probability that the mid-price moves up. The top pane shows the result obtained in 3 ways: blue and green are from Monte Carlo simulation and red from diffusion approximation. From the volumes at best bid/ask shown in the bottom pane, we can see the mid-price doesn't change during the whole period and it moves down at the end. The blue curve discovers this about 30 events earlier than the other two.

This is a simple trading strategy based on the probability calculation. Assuming we can freely long or short 1 share each time, we execute a long (short) whenever the probability is bigger than 0.9 (smaller than 0.1). Whenever we are at a long (short) position and the prediction is down (up), we clear our position. Green curve is the holdings throughout the day and blue is the cash value every time we hold no position. The prediction is made by Monte Carlo simulation (top) and diffusion approximation (bottom).

Cython Parallel Computing

The Monte Carlo simulations are very time consuming. All experiments are done using 80-core parallel computing in the Research Computing Center of FSU. By modifying Python code to Cython we obtain about x100 speed increase to get near C-performance. The single-core performance of MATLAB/Python/Cython/C is compared and the scalability of Cython and C with MPI is studied for the parallel computing.

References

[1] A. G. Hawkes. Spectra of Some Self-Exciting and Mutually Exciting Point Processes *Biometrika*, Vol. 58, No. 1. (1971), pp. 83-90

[2] I. Rubin. Regular Point Processes and Their Detection *IEEE Trans. Inform. Theory*, Vol. 18, No. 5. (1972), pp. 547-557

[3] Y. Ogata. On Lewis' Simulation Method for Point Processes *IEEE Trans. Inform. Theory*, Vol. 27, No. 1. (1981), pp. 23-31

[4] T. Yang, L. Zhu. A reduced-form model for level-1 limit order books arXiv:1508.07891v3 [q-fin.TR]