High-frequency data and limit order books

—

Lecture 5 - March 25th, 2025

Ioane Muni Toke

MICS Laboratory, FiQuant Team
CentraleSupélec – Université Paris-Saclay, France

U. Paris-Saclay CentraleSupelec cursus Ingénieur 3A Mathématiques et Data Science
IP Paris ENSAE M2 Statistics, Finance and Actuarial Science
U. Paris-Saclay Evry M2 Quantative Finance

Table of contents

Table of contents

Realized variance I

- Consider a (driftless) stochastic volatility model for a (log-)price p_t, $t \in [0, T]$:

$$p(t) = \int_0^t \sigma_s \, dB_s.$$

- The variance integrated over the time interval $[0, T]$ is

$$IV = [p]_T = \int_0^T \sigma_s^2 \, ds.$$

- For a given sampling period τ, we naturally define the realized variance estimator:

$$RV(\tau) = \sum_{i=1}^{\lfloor T/\tau \rfloor} (p(i\tau) - p((i-1)\tau))^2.$$

Realized variance II

- $RV(\tau)$ is a consistent estimator of IV :

$$RV(\tau) \xrightarrow[\tau \to 0]{} IV$$

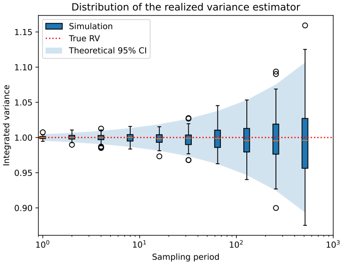
in probability.

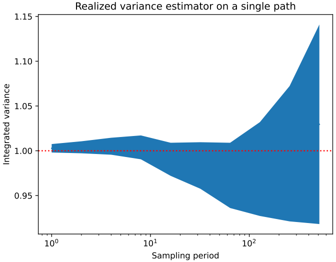
- The estimator $RV(\tau)$ is asymptotically Gaussian [3]:

$$\frac{RV(\tau) - IV}{\sqrt{\frac{2}{3}\sum_{i=1}^{\lfloor T/\tau \rfloor}(p(i\tau) - p((i-1)\tau))^4}} \xrightarrow[\tau \to 0]{\mathcal{L}} \mathcal{N}(0,1).$$

This gives us a measure of the estimation error for the realized variance (volatility).

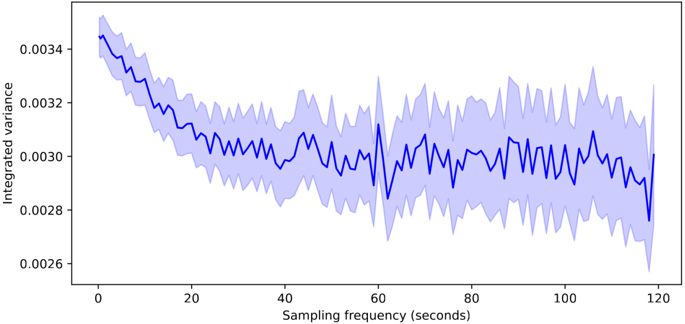
- Right column : illustration on a Brownian motion.





Volatility signature plot

- However on financial markets $RV(\tau)$ is empirically increasing when τ decreases. This is known as the volatility signature plot. Example on a French stock (2017 data):



- RV estimation on high-frequency financial data in contradiction with the Brownian paradigm: "microstructure noise". (Econometrics literature available, not discussed here.)

A Hawkes model for the signature plot I

- [1] proposes a model of Hawkes-driven upwards and downwards price jumps.
- Price is defined as $p(t) = N^1(t) - N^2(t)$, with $N^i, i \in \{1, 2\}$, Hawkes processes such that

$$\begin{cases} \lambda^1(t) = \lambda_0 + \int_{-\infty}^{t} \alpha e^{-\beta(t-s)} dN_s^2, \\ \lambda^2(t) = \lambda_0 + \int_{-\infty}^{t} \alpha e^{-\beta(t-s)} dN_s^1. \end{cases}$$

- No self-excitation of upward (resp. downward) jumps on following upward (resp. downward) jumps
- Only cross-excitation terms are kept, enforcing the mean-reversion empirically observed on the price p
- Cross-excitation is set to be symmetric.

A Hawkes model for the signature plot II

- The volatility signature plot plots the (normalized) realized variance as a function of the sampling period:

$$\hat{C}(\tau) = \frac{1}{T} \sum_{i}^{T/\tau} \left(\hat{p}(i\tau) - \hat{p}((i-1)\tau)\right)^2 .$$

- Mean signature plot in a stationary regime:

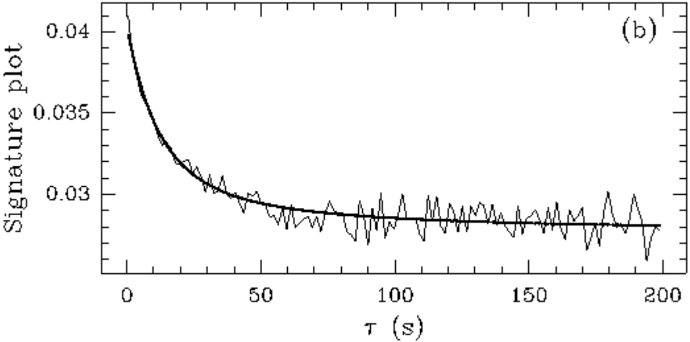
$$C(\tau) = \mathbf{E}[\hat{C}(\tau)] = \frac{1}{\tau} \mathbf{E}\left[p(\tau)^2\right]$$

- In the case of the above mean-reversion Hawkes model:

$$C(\tau) = \Lambda \left(\kappa^2 + (1-\kappa^2) \frac{1 - e^{-\gamma\tau}}{\gamma\tau} \right),$$

with $\Lambda = \frac{2\lambda_0}{1-\alpha/\beta}$, $\kappa = \frac{1}{1+\alpha/\beta}$, and $\gamma = \alpha + \beta$.

A Hawkes model for the signature plot III



Source: [1]

- Empirical data (Bund futures contracts 2009). Data: last bid price only (no bid-ask bounce).
- Signature plot reproduced by a Hawkes price model (Warning: result is not a MLE fit, but fit of the signature plot)

Table of contents

Correlation estimation with high-frequency data

- High-frequency financial data is non-synchronous: trades on stock 1 and trades on stock 2 do not occur at the same time.
- Estimation can be done via synchronous sampling, i.e. taking the last observed trade price at each sampling time, at the risk of introducing biases.
- If two continuous martingales P_1 and P_2 are observed on $[0, T]$ at asynchronous random times, the standard realized covariance estimator using the last observation is a biased estimator of $\mathbf{E}[\langle P_1, P_2 \rangle_T]$ (See [11]).
- A few estimators have been proposed to directly use asynchronous data.

The Hayashi-Yoshida estimator I

- Proposed in [11].
- Consider two processes ($k = 1, 2$):

$$dP_k(t) = \mu_k(t)dt + \sigma_k(t)dW_k(t)$$

with $(W_k(t))_t$ Brownian motions and $d\langle W_1, W_2\rangle_t = \rho(t)dt$, for some deterministic functions ρ, σ_1, and σ_2.

- Assume that we observe the process k on $[0, T]$ with at random times $T^{k,i}$. Let $I^i = (T^{1,i}, T^{1,i+1})$ and $J^i = (T^{2,i}, T^{2,i+1})$. Let n be the total number of observations.
- Let $\Delta P_1(I^i)$ be the increment of P on I^i (and define similarly $\Delta P_2(J^j)$). The Hayashi-Yoshida covariance estimator is

$$U_n = \sum_{i,j} \Delta P_1(I^i)\Delta P_2(J^j)\mathbf{1}_{I^i \cap J^j \neq \emptyset}.$$

The Hayashi-Yoshida estimator II

- The estimator is quite easy to implement.
- The estimator is consistent [11]:

$$U_n \to \int_0^T \sigma_1(t)\sigma_2(t)\rho(t)dt \text{ in probability when } n \to +\infty.$$

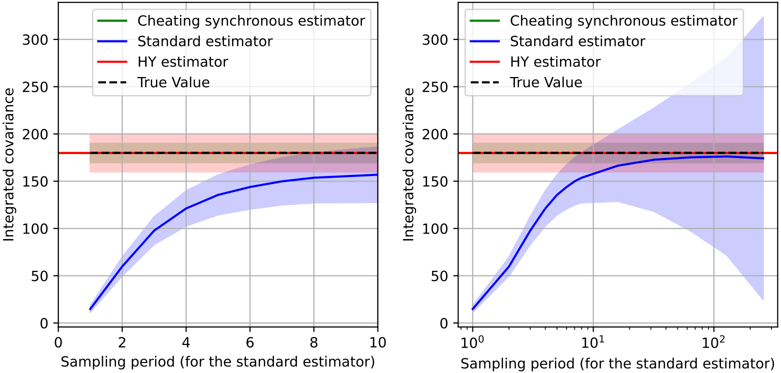
- The estimator is asymptotically Gaussian [10].
- In the case of constant ρ, σ_1 and σ_2, the correlation estimator is also consistent:

$$\frac{\sum_{i,j} \Delta P_1(I^i)\Delta P_2(J^j)\mathbf{1}_{I^i\cap J^j\neq\emptyset}}{\sqrt{\sum_i \Delta P_1(I^i)^2}\sqrt{\sum_j \Delta P_2(J^j)^2}} \to \rho \text{ in probability when } n \to +\infty.$$

(but careful, no asymptotic theory and may be numerically unstable)

The Hayashi-Yoshida estimator III

- An illustration on Brownian motions with Poisson sampling:



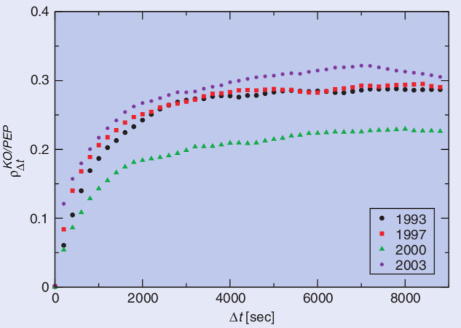
- On a similar subject, see also the Fourier (Malliavin-Mancino) estimator [14].

The Epps effect

- [5]: correlation measured on financial assets decreases when the sampling frequency increases (1979). Still observed on modern markets (e.g., [20], right panel below).

1. Correlations of Changes in Log Price for Four Stocks During Intervals of 10 Minutes to Three Days

	Pairs of Stocks					
Interval	*AMC-Chrysler*	*AMC-Ford*	*AMC-GM*	*Chrysler-Ford*	*Chrysler-GM*	*Ford-GM*
10 minutes	001	.009	– 009	– 014	007	055
20 minutes	009	018	011	017	026	118
40 minutes	.006	012	.014	041	040	197
One hour	–.043	.057	064	.023	065	.294
Two hours	029	.060	094	.112	129	383
Three hours	031	.158	111	361	518	519
One day	–.067	.170	078	342	442	571
Two days	– 020	223	186	336	449	572
Three days	– 098	203	.100	334	542	645



- Asynchronicity (cf. above), lead-lag ?

The Epps effect in a Hawkes model I

- [1] investigates the Epps effect using a bivariate version of the previous model:

$$\begin{cases} p_1(t) & = N^1(t) - N^2(t), \\ p_2(t) & = N^3(t) - N^4(t), \end{cases}$$

in which $\mathbf{N} = (N^i)_{i=1,\dots 4}$ is a 4D-Hawkes process with intensity:

$$\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_0 + \int_0^t \begin{pmatrix} 0 & \phi^{12} & \phi^{13} & 0 \\ \phi^{12} & 0 & 0 & \phi^{13} \\ \phi^{31} & 0 & 0 & \phi^{34} \\ 0 & \phi^{31} & \phi^{34} & 0 \end{pmatrix} (t-s) d\mathbf{N}_s, \tag{1}$$

where $\phi^{ij}(t-s) = \alpha^{ij} e^{-\beta^{ij}(t-s)}$.

The Epps effect in a Hawkes model II

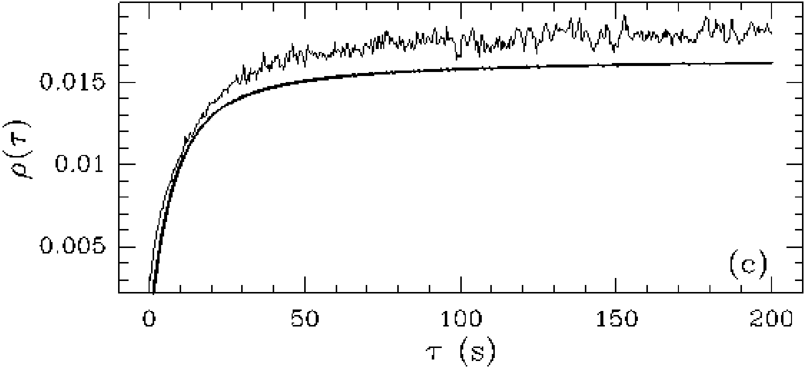
- Self-exciting terms are ruled out, $\phi^{ii} = 0$ for all i.
- Symmetric mean-reversion for each price process p_1 and p_2 ($\phi^{12} = \phi^{21}$, $\phi^{34} = \phi^{43}$).
- Only positive influence between prices p_1 and p_2 ($\phi^{14} = \phi^{23} = \phi^{32} = \phi^{41} = 0$).
- One price may have on stronger influence on the other, but this influence is symmetric w.r.t. the direction of the move ($\phi^{24} = \phi^{13}, \phi^{42} = \phi^{31}$).
- Explicit form of the correlation coefficient available

$$\rho(\tau) = \mathrm{Corr}\left(p_1(t+\tau) - p_1(t), p_2(t+\tau) - p_2(t)\right),$$

although quite cumbersome to write (see [1, Proposition 3.1]).

The Epps effect in a Hawkes model III

- Hawkes model in good agreement with the Epps effect. Bund/Bobl example:



Source: [1]

Table of contents

Reminder: Branching representation of a Hawkes process

Proposition (Branching representation of a Hawkes process)

Let $T > 0$. Let $\lambda_0 > 0$ and $\nu : [0, \infty) \to \mathbb{R}_+$ a deterministic function. Consider the following procedure:

1. *Generate a time-homogeneous Poisson process $\{t_i^{(0)}\}$ with intensity λ_0 on $[0, T]$.*
2. *For each $t_i^{(0)}$, generate a non-homogeneous Poisson process $\{t_i^{(1)}\}$ with deterministic intensity $t \mapsto \nu(t - t_i^{(0)})$ on $[t_i^{(0)}, T]$.*
3. *Repeat operation 2. for each point of all the sets $\{t_i^{(1)}\}$, then for each point of all the sets $\{t_i^{(2)}\}$, etc. until there is no new point generated in $[0, T]$.*

Then the ordered set of all the points generated in $[0, T]$ is a Hawkes process N on $[0, T]$ with intensity $\lambda_t = \lambda_0 + \int_0^t \nu(t - s)\, dN_s$.

Branching ratio, endogeneity

- Each (parent) point generates an average of $\int_0^{+\infty} \nu(t)\, dt = \|\nu\|_1$ (children) points. $\|\nu\|_1$ is called the branching ratio of the branching process.
- The average intensity is $\overline{\lambda} = \frac{\lambda_0}{1 - \|\nu\|_1}$. The ratio of endogeneous events in a stationary sample (endogeneity ratio ρ) is thus

$$\rho = 1 - \frac{\lambda_0}{\overline{\lambda}} = \|\nu\|_1.$$

- In other words, $\mathbf{E}[dN_t^{exo}] = \lambda_0\, dt$ and $\mathbf{E}[dN_t^{endo}] = \overline{\lambda}\|\nu\|_1\, dt$.
- In a financial context, endogenous vs. exogenous trading. Careful calibration needed. Are financial markets in a critical regime ? See eg. [8, 9, 7].

Causality

- Let us consider a stationary D-dimensional Hawkes process with constant baseline intensities $\mu = (\mu^i)_{i=1,\dots,D}$ and kernel matrix $\Phi = (\Phi_{ij})_{i,j=1,\dots,D}$.
- Let N_t^{i0} be the counting process of events of type i generated exogenously. Then

$$\mathbf{E}[dN_t^{i0}] = \mu^i \, dt.$$

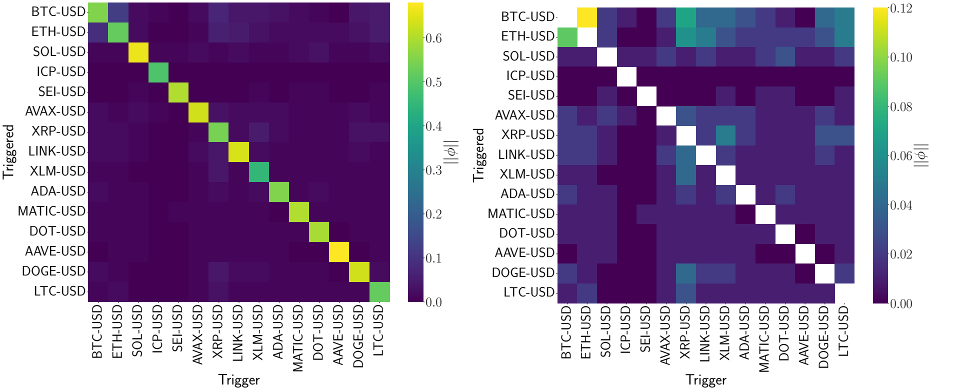
- Let N_t^{ij} be the counting process of events of type i with a direct ancestor of type j.

$$\mathbf{E}[dN_t^{ij}] = \|\Phi_{ij}\|_1 \Lambda_j \, dt.$$

  where $\Lambda = (\mathbf{I} - \|\Phi\|_1)^{-1}\mu$ is the vector of average intensities.
- This provides a causality measure that can be used on financial markets.

An example of analysis on cryptocurrency pairs

- Endogeneity matrices on trade times of 15 cryptocurrency pairs on a centralized exchange.



Source : [6]

Table of contents

A Hawkes model for aggressive orders I

- [12] is an early attempt at modeling many types of LOB events with a multidimensional Hawkes process. Events are divided in to ten categories *(right)*.

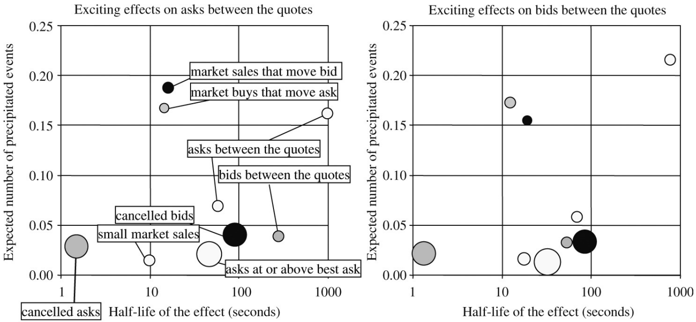
Type	Description	Aggressiveness
1	Market order that moves the ask	Yes
2	Market order that moves the bid	Yes
3	Limit order that moves the ask	Yes
4	Limit order that moves the bid	Yes
5	Market order that doesn't move the ask	No
6	Market order that doesn't move the bid	No
7	Limit order that doesn't move the ask	No
8	Limit order that doesn't move the bid	No
9	Cancellation at ask	No
10	Cancellation at bid	No

- A 4-dimensional Hawkes processes for aggressive orders is proposed ($m = 1, \ldots, 4$):

$$\lambda^m(t) = \lambda_0(t) + \sum_{n=1}^{10} \int_0^t \alpha_{mn} e^{-\beta_{mn}(t-u)} dN_u^n.$$

A Hawkes model for aggressive orders II

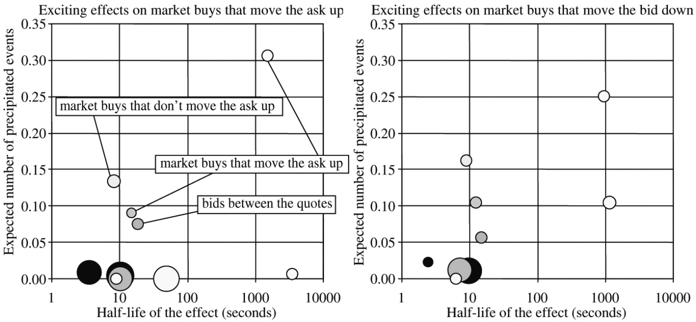
- Estimates for aggressive limit orders



- Reported findings:
  - aggressive limit orders are firstly influenced by aggressive market orders: this is an evidence of some resiliency or market making in the order book;
  - aggressive limit orders are secondly influenced by aggressive limit orders.

A Hawkes model for aggressive orders III

- Estimates for aggressive market orders



- Reported findings:
  - aggressive market orders are firstly influenced by market orders (aggressive or not);
  - aggressive market orders are secondly influenced by aggressive limit orders, which might be labeled a rush to liquidity.

A general Hawkes LOB model – 0,1 event description for the first limits I

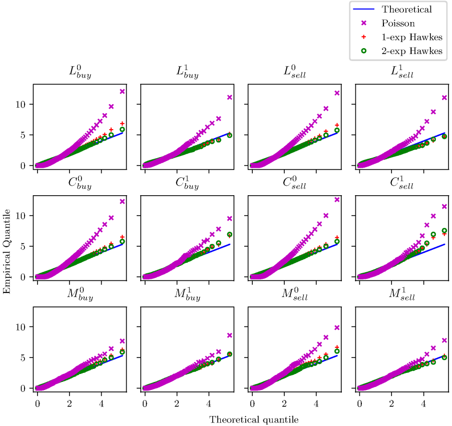
- [13] proposes a large Hawkes model with all events happening at the best quotes.
- Three types of orders (M,L,C), two aggressivities (0/non-aggressive,1/aggressive) and two sides (bid/ask) yield 12 event types: $E = \{L^0_{buy}, L^0_{sell}, C^0_{buy}, C^0_{sell}, M^0_{buy}, M^0_{sell}, L^1_{buy}, L^1_{sell}, C^1_{buy}, C^1_{sell}, M^1_{buy}, M^1_{sell}\}$.
- 12-dimensional Hawkes process $N(t) = (N_{L^0_{buy}}(t), \dots, N_{M^1_{sell}}(t))$, with intensity process $(\lambda_{L^0_{buy}}(t), \dots, \lambda_{M^1_{sell}}(t))$.
- Variants tested: 1 or 2 exponential kernels:
  - $\phi_{ij}(t) = \alpha_{ij} \exp(-\beta_{ij} t)$
  - $\phi_{ij}(t) = \sum_{p=1}^{2} \alpha_{ijp} \exp(-\beta_{ijp} t)$

A general Hawkes LOB model – 0,1 event description for the first limits II

- LOB is not modelled *per se* (no tracking of limits)
- Simplifying assumption : all bid and ask price movements are equal to one tick (ticksize $\eta > 0$); OK for large tick stocks ; in the paper sample: 1.08 ticks per jump in average.
- (A proxy for the) mid-price dynamics is obtained as a by-product of event arrivals:

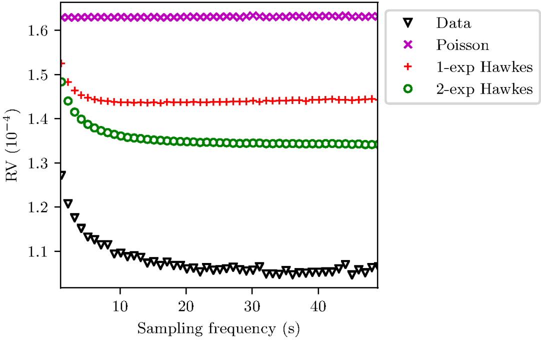
$$P(t) = P(0) + \left(\sum_{e \in E_{up}} N_e(t) - \sum_{e' \in E_{down}} N_{e'}(t) \right) \times \frac{\Delta P}{2}, \quad t > 0$$

A general Hawkes LOB model – Goodness-of-fit



- Transformed durations should be exponentially distributed (see previous lecture)
- Q-Q plot goodness of fit tests of order book models
- Poisson model globally fails to capture the distributional properties of interval times.
- 1- and 2-exponential Hawkes models: similar performances, except for orders of type 0
- 1- and 2-exponential Hawkes models: close to one another, but not very close to data

A general Hawkes LOB model – Signature plot



- Mean signature plot.
- Expected: signature plot of the Poisson model is flat
- 1-exponential and 2-exponential Hawkes: the realized volatility decreases when the sampling interval increases, but the long-term volatility level is too high compared to the data.

Inconsistencies between data and the linear Hawkes model

- Conditional probabilities of pairs $(X|Y)$ (probability of observing an event of type X given that the previous event is of type Y):

Pair	P_{simu}	P_{real}
$C^1_{buy}\|L^0_{buy}$	0.402	0.048
$L^1_{buy}\|L^1_{buy}$	1.628	0.141
$L^1_{sell}\|L^1_{buy}$	1.288	0.171
$M^0_{sell}\|C^1_{buy}$	0.545	0.068
$C^1_{buy}\|C^1_{buy}$	0.548	0.072
$M^1_{sell}\|C^1_{buy}$	0.854	0.037

- $C^1_{buy}|L^0_{buy}$: needs two cancellations within the same microsecond ; $L^1_{buy}|L^1_{buy}$ and $L^1_{sell}|L^1_{buy}$: rare because spread constraints ; orders following a C^1_{buy}: participants not willing to widen the spread.
- Further evidence : medians of L_1 norm of kernels $\phi_{\cdot C^1_{buy}}$ in the 2-exponential model.

L^0_{buy}	L^0_{sell}	C^0_{buy}	C^0_{sell}	M^0_{buy}	M^0_{sell}
0.1563	0.2357	0.9392	0.0914	0	0
L^1_{buy}	L^1_{sell}	C^1_{buy}	C^1_{sell}	M^1_{buy}	M^1_{sell}
0.3845	0.1607	0	0	0.0013	0

Inhibition effects cannot be modelled in a framework with excitation only.

Inhibition in LOB models – Nonlinear Hawkes processes

- General non-linear Hawkes model:

$$\lambda(t) = \varphi\left(\lambda_0 + \int_0^t \Phi(t-s)\, dN_s\right).$$

- [13] modifies the previous model by including inhibition for the pairs identified above with a negative 2-exponential kernels:

$$\phi_{mn}(t) = \sum_{p=1}^{2} -\alpha_{mnp} \exp(-\beta_{mnp} t),$$

  where the α's and β's are non-negative real numbers, and the non-linear function $\varphi(x) = \max(x, 0)$ to keep the intensity positive.
- Some general results on non-linear Hawkes processes in e.g., [4, 22] (outside the scope of this course).

Nonlinear Hawkes processes – Conditional probabilities and inhibition kernels

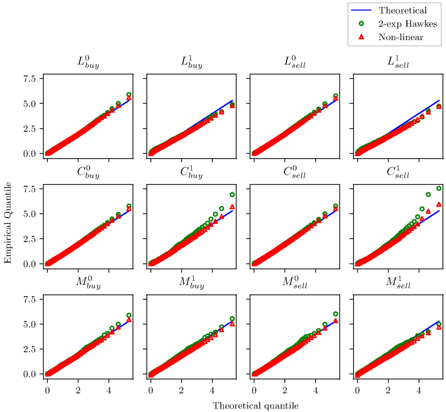
- Pairs previously identified as badly represented by the linear models:

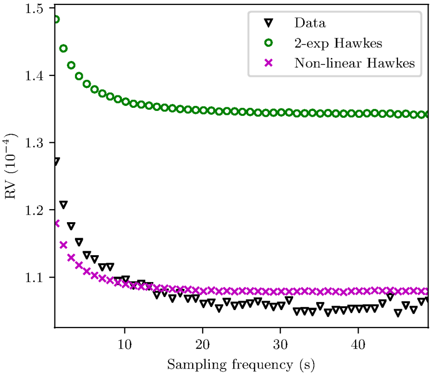
Pair	P_{simu}	P_{simu}^{NL}	P_{real}
$C_{buy}^1 \vert L_{buy}^0$	0.402	0.105	0.048
$L_{buy}^1 \vert L_{buy}^1$	1.628	0.254	0.141
$L_{sell}^1 \vert L_{buy}^1$	1.288	0.235	0.171
$M_{sell}^0 \vert C_{buy}^1$	0.545	0.135	0.068
$C_{buy}^1 \vert C_{buy}^1$	0.548	0.163	0.072
$M_{sell}^1 \vert C_{buy}^1$	0.854	0.108	0.037

- Medians of L_1 norm of kernels $\phi_{\cdot C_{buy}^1}$ in non-linear model for those that were 0 in 2-exponential model:

M_{buy}^0	M_{sell}^0	C_{buy}^1	C_{sell}^1	M_{sell}^1
-0.0319	-0.1593	-0.0541	-0.1439	-0.1908

Nonlinear Hawkes processes – Goodness-of-fit and signature plots





Inhibition effects exhibits satisfying goodness-of-fit and better reproduction of the volatility.

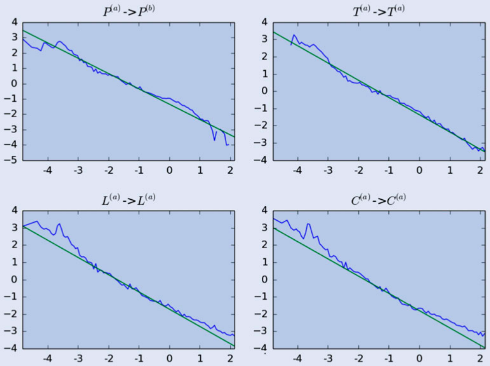
Non-parametric kernels – Estimation on futures data

- In [2], a Hawkes model with non-parametric kernels is fitted to the first limits with the following events $N_t = (P^{(a)}, P^{(b)}, T^{(a)}, T^{(b)}, L^{(a)}, L^{(b)}, C^{(a)}, C^{(b)})$(mid-price moves, then trades, limit orders and cancellations that do not move the price)

- Data: Level-I data on BUND and DAX futures.
- Slowly decreasing kernels are found for many interactions (log-log-scale).

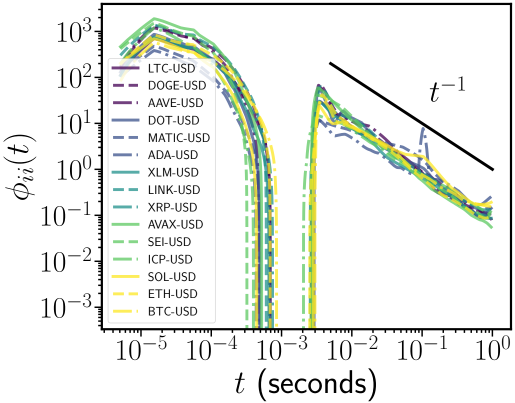
Exponential kernels are usually not found in empirical analysis of microstructure events.

- Multiple exponential kernels is a workaround for longer dependencies.



Source: [2]

Non-parametric kernels – Estimation on cryptocurrencies



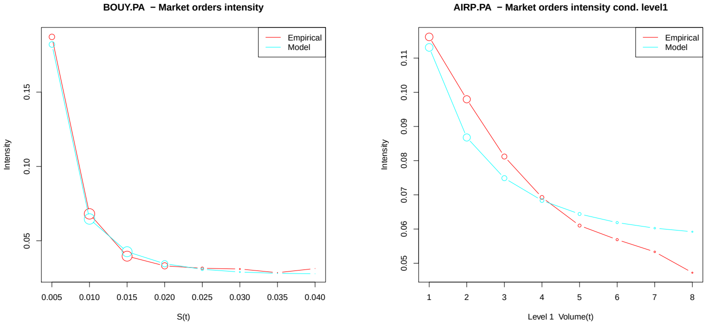
- In [6] a 15-dimensional non-parametric Hawkes model is fitted to cryptocurrency trade data.
- Long memory observable on cryptocurrencies as well.
- Non-parametric estimation reveals complex structure of cross- and self-excitation in fitted Hawkes models.

Source : [6]

Table of contents

State-dependent intensities

Empirical intensities of submission of market orders are state-dependent in e.g., spread *(left)* and the size of the best queue *(right)*.



Source: [17]

- Light-blue curve : fit of a parametric model
$\lambda^M(t; S(t), q_1(t)) = \exp\left[\beta_0 + \beta_1 \ln(S(t)) + \beta_{11}[\ln(S(t))]^2 + \beta_2 \ln(1+q_1(t)) + \beta_{22}[\ln(1+q_1(t))]^2 + \beta_{12} \ln(S(t)) \ln(1+q_1(t))\right].$

A first state-dependent Hawkes models I

- [15] proposes a Hawkes extension in which the kernels depend on a state vector (e.g., representing the state of the LOB).
- Let $\mathcal{E}$ be a set of point processes with pooled event times $(T_n)_{n\in\mathbb{N}}$. Let E_n be the type of event observed at time T_n. The intensity process for the process $e \in \mathcal{E}$ is:

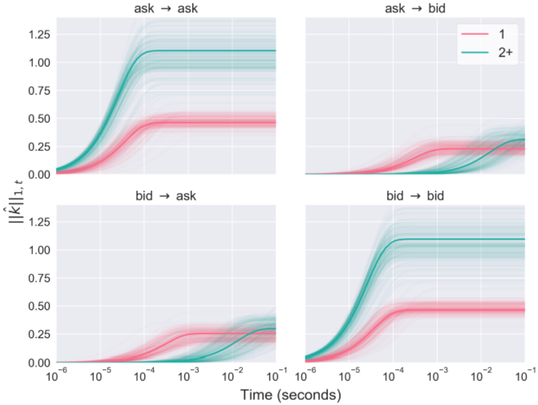
$$\lambda_e(t) = \nu_e + \sum_{e'\in\mathcal{E}} \int_{[0,t)} k_{ee'}(t-s, X_s)\, dN_{e'}(s)$$

  and the state process X is piecewise constant with jumps $(T_n)_{n\in\mathbb{N}}$ s.t.

$$\mathbf{P}(X(T_n) = x | E_n, \mathcal{F}_{T_n-}) = \Phi_{E_n}(X(T_n-), x),$$

  where the Φ_{E_n}'s are transition probability matrices.
- Model is applied to imbalance/spread models for LOB events.

A first state-dependent Hawkes models II



(a) Models (state variable: bid–ask spread).

- Example of results for upward/downward events ($|\mathcal{E}| = 2$) and state=spread *(left)*
- State-dependency obvious.
- Potential parsimony problems.

Source : [15]

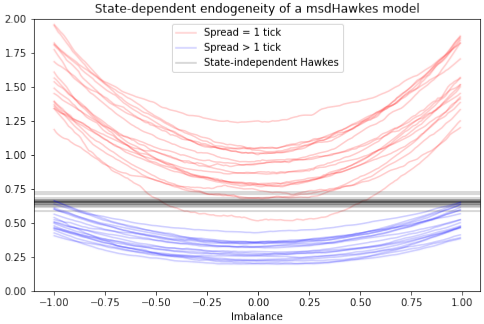
Another state-dependent Hawkes model I

- [19] extends the standard Hawkes model with a state-dependent multiplicative factor. Let $d_e \in \mathbb{N}^*$. Let $(N_t)_{t \geqslant 0}$ be a d_e-dimensional counting process with stochastic intensity λ.
- State representation : Let $d_x \in \mathbb{N}$. Let $(X_t)_{t \geqslant 0}$ be an adapted, piecewise constant, left-continuous stochastic process, with values in $[-1, 1]^{d_x}$. $(X_t)_{t \geqslant 0}$ denotes the observable state space.
- Hawkes process with multiplicative state-dependent factor (msdHawkes) : point process N with an intensity $\lambda = (\lambda^1, \ldots, \lambda^{d_e})$ of the form:

$$\lambda^e(t) = \left(\nu + \int_{]0,t[} \phi(t-s) \cdot dN_s\right) \exp\left(\langle \theta^e, X_t \rangle\right), \tag{2}$$

  where $(\theta^1, \ldots, \theta^{d_e}) \in (\mathbb{R}^{d_x})^{d_e}$ are the state coefficients.
- See also [21] for an analogous queue-size-dependent Hawkes order flow model.

Another state-dependent Hawkes model II



Importance of state-dependence, but potential complex interplay of regimes outside of standard theory.

A state-dependent spread Hawkes model I

- If $S_t^{+/-k}$ counts the number of positive/negative jumps of the spread of size k ticks and assuming the maximum jump size is K, the spread process is

$$S_t = S_0 + \sum_{k=1}^{K} k\, S_t^{+k} - \sum_{k=1}^{K} k\, S_t^{-k}.$$

- [18] propose a state-dependent spread Hawkes model:

$$\lambda_t^e = f^e(S_{t-}) \left[\mu^e + \sum_{e' \in \mathcal{E}} \int_0^t \Phi_{ee'}(t-s)\, dS_s^e \right],$$

  where $\mathcal{E} = \{-K, \dots, -1, 1 \dots, K\}$ is the space of jump sizes.
- The state-dependent function is one way to ensure that the spread remains positive: $f^{-k}(n) = 0$ if $n \leq k$.
- Normalization necessary (all quantities defined up to a constant because of f)

A state-dependent spread Hawkes model II

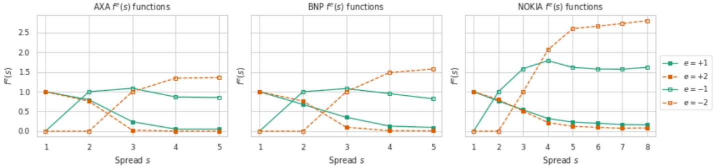


Figure 3: Estimations of the $\{f^e(s)\}_{e\in\mathcal{E}}$ functions for AXA (left), BNP (middle), NOKIA (right),

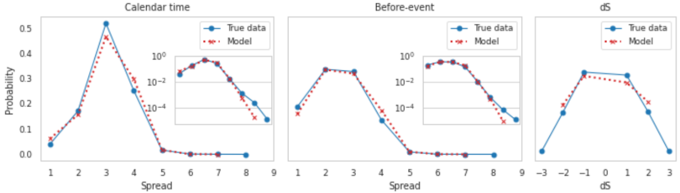


Figure 6: Spread distribution, comparison between true data and the the data obtained through simulation for AXA. The left-hand figure is the Calendar time distributions, the middle figure is the event time distributions, and the right-hand figure is the distributions of spread jumps size.

- The state-dependent factor is crucial for accurate reproduction of empirical data.

Source : [18]

An intensity ratio model I

- One may also focus on the state-dependence by assuming a baseline intensity $\lambda_0(t)$, neither observable nor specified as a function of observables and parameters, common to all coordinates. Intensity $\lambda^i(t)$ of the process $(N^i_t)_{t\geq 0}, i \in \mathbb{I}$, is written.

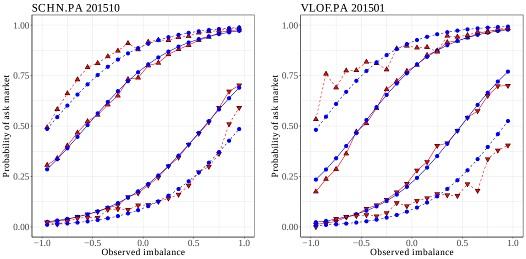
$$\lambda^i(t,\vartheta) = \lambda_0(t)\lambda^i(X(t)),$$

  where $(X(t))_{t\geq 0}$ is an observable covariate process
- In the case $\lambda^i(X(t)) = \exp\left(\sum_{j\in\mathbb{J}} \vartheta^i_j X_j(t)\right)$ for a parameter vector $\vartheta = (\vartheta^i_j)_{i\in\mathbb{I},j\in\mathbb{J}}$, MLE estimation of the *relative* dependencies available via the intensities ratios ([16]):

$$r_i(t,\theta) = \frac{\lambda^i(t,\vartheta)}{\sum_{i'\in\mathbb{I}}\lambda^{i'}(t,\vartheta)} = \left[1 + \sum_{i'\in\mathbb{I}\setminus\{i\}} \exp\left(\sum_{j\in\mathbb{J}}(\vartheta^{i'}_j - \vartheta^i_j)X_j(t)\right)\right]^{-1}.$$

- $\lambda^i(X(t))$ can also be flexibly represented with a neural network.

An intensity ratio model II



Examples of fitted order flows probabilities in a ratio model.

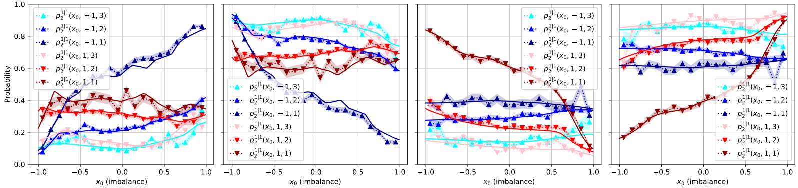


Table of contents

Lab 4 - Questions

For this lab you need to add a new set of data to your environment (while keeping the hierarchical file structure `./Data/{stock}/{files.csv.gz}`).

1. **Brownian motions with Poisson sampling.** Take your results from Lab 2 Question 2 and add the Hayashi-Yoshida estimator to the plot.
2. **Signature plots and Epps effect.** On your dataset, estimate the integrated realized variances for each stock and the correlation between the two stocks using the standard estimators and the Hayashi-Yoshida estimators. Comment.
3. **Hawkes fit on signature plots.** Fit the single stock Hawkes model of [1] on the signature plot for each stock of your dataset. Comment. Simulate the model with the calibrated parameter values and plot the empirical, fitted and simulated signature plots.

References I

[1] E. Bacry, S. Delattre, M. Hoffmann, and J. F Muzy. "Modeling microstructure noise with mutually exciting point processes". In: *Quantitative Finance* 13.1 (2013), pp. 65–77.

[2] E. Bacry, T. Jaisson, and J.-F. Muzy. "Estimation of slowly decreasing Hawkes kernels: application to high-frequency order book dynamics". In: *Quantitative Finance* 16.8 (2016), pp. 1179–1201.

[3] O. E. Barndorff-Nielsen and N. Shephard. "Econometric analysis of realized volatility and its use in estimating stochastic volatility models". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 64.2 (2002), pp. 253–280.

[4] P. Brémaud. "Stability of nonlinear Hawkes processes". In: *The Annals of Probability* 24.3 (1996), pp. 1563–1588.

[5] T. W. Epps. "Comovements in stock prices in the very short run". In: *Journal of the American Statistical Association* 74.366a (1979), pp. 291–298.

References II

[6] T. Fabre and I. Muni Toke. "Neural Hawkes: Non-Parametric Estimation in High Dimension and Causality Analysis in Cryptocurrency Markets". In: *Quantitative Finance* (2025). In press.

[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". In: *Quantitative Finance* 15.8 (2015), pp. 1293–1314.

[8] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud. "Critical reflexivity in financial markets: a Hawkes process analysis". In: *The European Physical Journal B* 86.10 (2013), pp. 1–9.

[9] S. J. Hardiman and J.-P. Bouchaud. "Branching-ratio approximation for the self-exciting Hawkes process". In: *Physical Review E* 90.6 (2014), p. 062807.

[10] T. Hayashi and N. Yoshida. "Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes". In: *Annals of the Institute of Statistical Mathematics* 60 (2008), pp. 367–406.

References III

[11] T. Hayashi and N. Yoshida. "On covariance estimation of non-synchronously observed diffusion processes". In: *Bernoulli* 11.2 (2005), pp. 359–379.

[12] J. Large. "Measuring the resiliency of an electronic limit order book". In: *Journal of Financial Markets* 10.1 (2007), pp. 1–25.

[13] X. Lu and F. Abergel. "High dimensional Hawkes processes for limit order books". In: *Quantitative Finance* 18.2 (2018), pp. 249–264.

[14] P. Malliavin and M. E. Mancino. "Fourier series method for measurement of multivariate volatilities". In: *Finance and Stochastics* 6 (2002), pp. 49–61.

[15] M. Morariu-Patrichi and M. S. Pakkanen. "State-dependent Hawkes processes and their application to limit order book modelling". In: *arXiv preprint arXiv:1809.08060* (2018).

[16] I. Muni Toke and N. Yoshida. "Analyzing order flows in limit order books with ratios of Cox-type intensities". In: *Quantitative Finance* 20.1 (2020), pp. 81–98.

References IV

[17] I. Muni Toke and N. Yoshida. "Modelling intensities of order flows in a limit order book". In: *Quantitative Finance* 17.5 (2017), pp. 683–701.

[18] R. Ruan, E. Bacry, and J.-F. Muzy. "The self-exciting nature of the bid-ask spread dynamics". In: *arXiv preprint arXiv:2303.02038* (2023).

[19] E. Sfendourakis and I. Muni Toke. "LOB modeling using Hawkes processes with a state-dependent factor". In: *Market microstructure and liquidity* (2023).

[20] B. Tóth and J. Kertész. "The Epps effect revisited". In: *Quantitative Finance* 9.7 (2009), pp. 793–802.

[21] P. Wu, M. Rambaldi, J.-F. Muzy, and E. Bacry. "Queue-reactive Hawkes models for the order flow". In: *arXiv preprint arXiv:1901.08938* (2019).

[22] L. Zhu. "Large deviations for Markovian nonlinear Hawkes processes". In: *The Annals of Applied Probability* 25.2 (2015), pp. 548–581.