Dynamique Haute-Fréquence des marchés financiers

Emmanuel Bacry

Directeur de recherche au CNRS
Professeur Associé
Responsable de l'Initiative Data Science
Centre de Mathématiques Appliquées
Ecole Polytechnique

emmanuel.bacry@polytechnique.fr
http://www.cmap.polytechnique.fr/~bacry

Introduction to Hawkes processes

- Point processes introduced by A.G.Hawkes in the 70's
- Flexible and versatile tool to investigate mutual and/or self interaction of dynamic flows
- Very sucessful in seismic ($>$ 1980)
- Rising popularity in finance ($>$ 2007)
- Rising popularity in machine learning (network, ...)

The 1-Dimensional Poisson process

- N_t : a jump process (jumps are all of size 1)
- λ_t : intensity ($\simeq$ density of jumps)

$$\lambda_{\mathbf{t}} = \mu$$

$\Longrightarrow$ **The inter-arrival times are independant**

The Poisson process

- N_t : jump process (jumps are all of size 1)
- λ_t : the intensity
- μ : 1-dimensional exogenous intensity

$$\lambda_{\mathbf{t}} = \mu$$

A Hawkes process

$\Longrightarrow$ Introducing (positive) correlation in the arrival flow

$\Longrightarrow$ "Auto-regressive" relation

$$\lambda_{\mathbf{t}} = \mu + \phi \star \mathbf{dN_t},$$

where by definition

$$\phi \star dN_t = \int_{-\infty}^{+\infty} \phi^{ik}(t-s)dN(s)$$

and $\phi(t)$: kernel function which is positive and causal (i.e., supported by R^+).

The Poisson process

- N_t : jump process (jumps are all of size 1)
- λ_t : the intensity
- μ : 1-dimensional exogenous intensity

$$\lambda_{\mathbf{t}} = \mu$$

A Hawkes process

$\Longrightarrow$ Introducing (positive) correlation in the arrival flow
$\Longrightarrow$ "Auto-regressive" relation

$$\lambda_{\mathbf{t}} = \mu + \phi \star \mathbf{dN_t},$$

where by definition

$$\phi \star dN_t = \int_{-\infty}^{+\infty} \phi^{ik}(t-s) dN(s)$$

and $\phi(t)$: kernel function which is positive and causal (i.e., supported by R^+).

Hawkes processes - general definition in dimension D

- N_t : a D-dimensional jump process (jumps are all of size 1)
- λ_t : D-dimensional stochastic intensity
- μ : D-dimensional exogenous intensity
- $\Phi(t)$: $D \times D$ square matrix of kernel functions $\Phi^{ij}(t)$ which are positive and causal (i.e., supported by R^+).
- Moreover $||\Phi^{ij}||_{L^1} < +\infty, \quad 1 \leq i,j \leq D$

"Auto-regressive" relation

$$\lambda_t = \mu + \Phi \star dN_t,$$

where by definition

$$(\Phi \star dN_t)^{ij} = \sum_{k=1}^{D} \int_{-\infty}^{+\infty} \Phi^{ik}(t-s) dN^k(s)$$

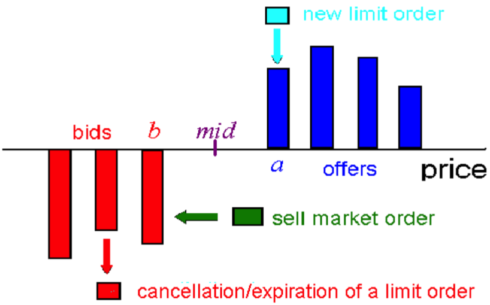
Hawkes processes : Challenges

- Flexible and versatile tool to investigate mutual and/or self interaction of dynamic flows
- Very sucessful in seismic (> 1980)
- Rising popularity in finance (> 2007)
- Rising popularity in machine learning (network, ...)
- **Many challenges** :
  - (non) parametric estimation,
  - high dimension,
  - Big data (scalability),
  - ...

High-frequency LOB dynamics

- Emergence of electronic markets $\Rightarrow$ Growing activity
- Limit Order Books (LOB) are at the heart of the price formation process
- **Understanding LOB high frequency dynamics has become fundamental**
    - For policy makers (regulations)
    - Market makers (liquidity)
    - Investors (reducing transaction cost, impact)
    - ...

Price formation process : the Limit Order Book (LOB)



- Limit order : offer to buy (sell) a specific quantity at a certain price
- Cancellation : withdrawal of a limit order
- Market order : order to buy (sell) at best available price

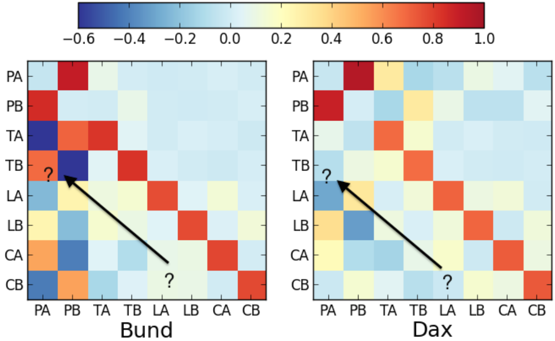
An 8-dimensionnal model for best bid/ask events

E.B. T.Jaisson and J.F. Muzy (2015)

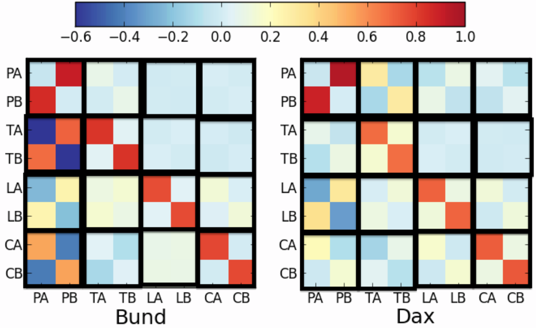
- **Database** :
  - Dax Futures (small tick size), Bund Futures (large tick size)
  - 1 year data : 06/2013-06/2014
  - **timestamp precision** $= \mathbf{1}\mu s$ (actual precision $\simeq 10\mu s$)
- **8-dimensional flow** :
  - PA (resp. PB) : upward (resp. downward) mid-price jumps
  - TA (resp. TB) : market orders at the best ask (resp. bid)
  - LA (resp. LB) : limit orders at the best ask (resp. bid)
  - CA (resp. CB) : cancel orders at the best ask (resp. bid)
- **More than 100 million events for one asset**
  - Multithreaded cluster (for one asset)
  - HDFS cluster (for many assets)
  - FPGA ?

Revealing High-frequency dynamics

Color coding of the self/mutual excitations/inhibitions

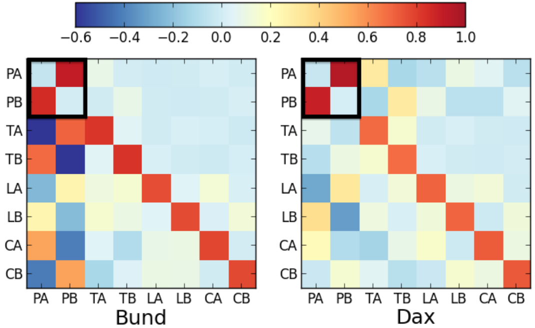


Mutual/Self excitations of flows



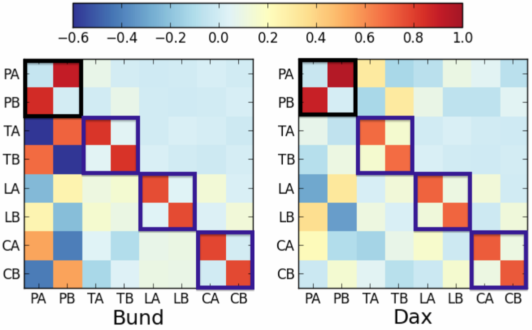
⟹ **Symmetry upward/downward and ask/bid**

Mutual/Self excitations of Downward/Upward Price flows



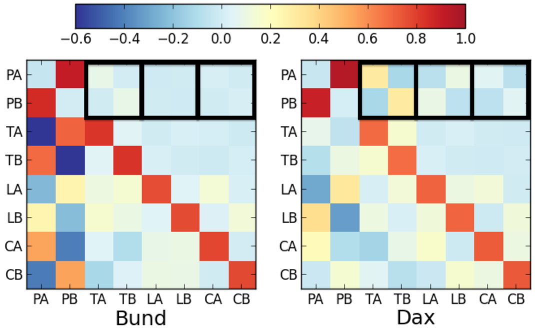
- "Anti-diagonal" shape in the price kernels
  $\Rightarrow$ **mean reversion of the price**

Mutual/Self excitations of Order flow



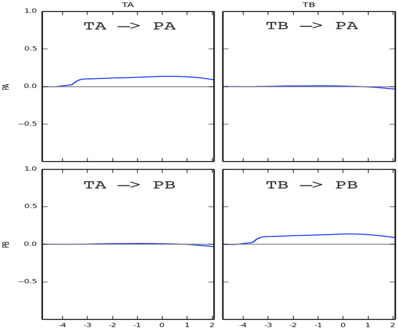
- “Anti-diagonal” shape in the price kernels
  $\Rightarrow$ **mean reversion of the price**
- “Diagonal” shape in the limit/cancel/trade kernels
  $\Rightarrow$ **splitting/herding**

Impact of the Order flows on the Price



- Trades : main source of impact (diagonal)
- Limits : contrariant
- Cancels : diagonal

Impact of the Market Order flows on the Price



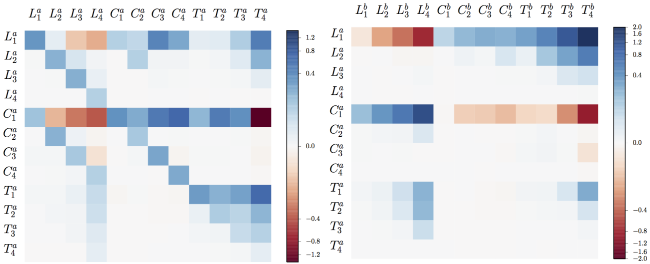
Cumulative kernels $\int_0^t \Phi^{T? \to P?}(s)ds$ as a function of $\log(t)$

- **Impact kernels $\Phi^{TA \to PA}$ and $\Phi^{TB \to PB}$ are very localized**
- Localization around "latency value" $\simeq 0.25ms$

Going to higher dimensions ...

Account for the volume in .... dimension 24 !

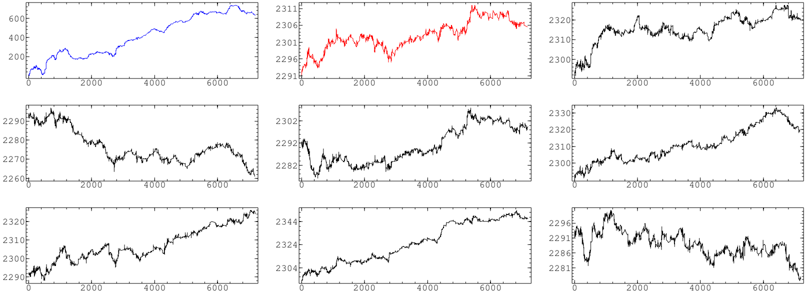
M.Rambaldi, E.B., F.Lillo (2016)



Ask→ Ask interaction (left) and Bid → Ask interaction (left)
(Only two quadrants shown)

Current work : "Labelled" EuroNext database, >1000 agents
⟶ Herding ? Instability tracking ? Information flow tracking ?

Replay of 2 hours of Eurostoxx mid-price from real trades



$TA_t - TB_t$: True cum. Trades on 3/08/2008 - [10am-12am]

$PA_t - PB_t$: True mid-price on 3/08/2008 between 10am and 12am

Simulation of the mid-price process $X = PA - PB$ given the real trades