Modelling assets volatilities and correlations

from microscopic dynamics to the diffusive regime

Emmanuel Bacry

Senior researcher, CNRS
Univ. Paris-Dauphine, PSL

PR[AI]RIE Institute

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

September 13, 2023

- N_t : a 1-dimensional Hawkes process
- λ_t : stochastic intensity
- μ : exogenous intensity
- $\phi(t)$: kernel function $\Phi^{ij}(t)$ which are positive and causal (i.e., supported by R^+).

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

with

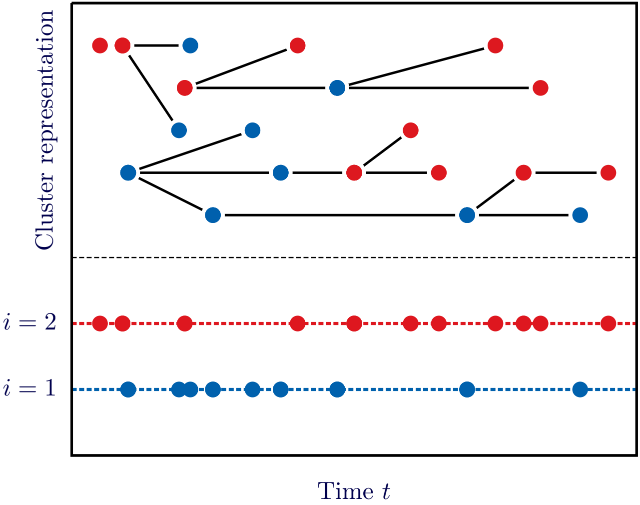
$$\phi \star dN_t = \int_{-\infty}^{+\infty} \phi(t-s) dN^k(s)$$

- N_t : a D-dimensional Hawkes process
- 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^+).

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

with

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


Cluster representation
$i = 2$
$i = 1$
Time t

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

For each component:

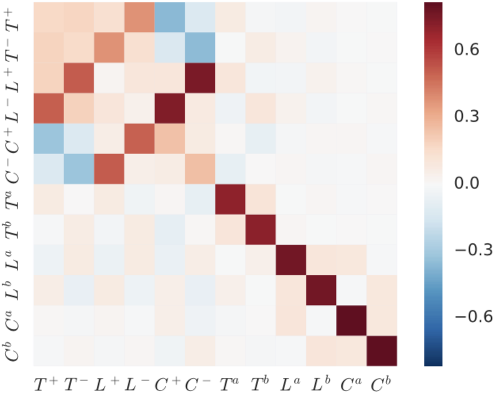
- $||\Phi||^{ij} = \|\phi^{ij}\|$ is the average number of type i event triggered directly by a type j event.
- $\Lambda^i = \mathbb{E}\left[\lambda(t)\right]$ = averaged number of events of type i
- μ^i is the immigrant intensity of type i.
- The Reaction matrix

$$\boldsymbol{R} = (\boldsymbol{I} - ||\Phi||)^{-1}$$

- R^{ij} = averaged number of events of type i triggered in total (directly and indirectly) by an event of type j
- $\Lambda^i = \mathbb{E}\left[\lambda_t^i\right] = \sum_j R^{ij}\mu_j$

(*Achab, Bacry, Muzy, Rambaldi, 2018*)

High-frequency financial data (DAX order book dynamics)
Estimation of 144 kernels



(Rambaldi, Bacry, Muzy, 2018)

N_t : order arrivals in the orderbook (each component of N_t represents a different order type)

Let δ_i be the mid-price change determined by an event/order of type i, then

$$\Delta_\tau P(t) \equiv P(t+\tau) - P(t) = \sum_{i \in M} \delta_i \int_t^{t+\tau} dN_s^i$$

And the volatility at time scale τ :

$$\sigma_\tau^2 = E(\Delta_\tau P^2) = \sum_{i,j \in M} \delta_i \delta_j \int_0^\tau \int_0^\tau E(dN_s^i dN_{s'}^j)$$

Putting together

- $\boldsymbol{R} = (\boldsymbol{I} - ||\Phi||)^{-1}$
- $\sigma_\tau^2 = \sum_{i,j \in M} \delta_i \delta_j \int_0^\tau \int_0^\tau E(dN_s^i dN_{s'}^j)$

One gets, after some calculations, for the diffusive volatility :

$$\frac{\sigma_\tau^2}{\tau} \xrightarrow[\tau \to \infty]{} \sum_m \Lambda^m \xi_m^2 = \sum_{m=1}^n \Lambda^m \left(\sum_{i \in M} \delta_i R^{im} \right)^2$$

where

$$\xi_m = \text{average volatility per event of type } m$$

We have a link from microscopic dynamics to the diffusive regime

- $N_{i,\alpha}(t)$ counting process associated with actions α of agent i.
- We will suppose that $i = 1, \ldots, M$ and $\alpha \in \mathcal{A} = \{P^+, P^-, L^a, L^b, C^a, C^b, T^a, T^b\}$ where
  - P^+ (P^-) orders that immediately move upward (downward) the mid-price;
  - T^a (T^b) aggressive orders at the best ask (bid) that do not move the price;
  - L^a (L^b) new limit orders that arrive at the best ask (bid);
  - C^a (C^b) cancel orders at the best ask (bid) that do not move the price;
- $\phi^{i,\alpha;j,\beta}$ = influence of order type β of agent j on order type α of agent i

Total number of interactions: $(M \times 8) \times (M \times 8)$
For $M = 15$ agents that's 14400 kernels to estimate !!

Such huge number of kernels is hard to handle.

$\downarrow$

Work hypothesis:

- Influence on agent i from agent j actions does not depend on j provided $j \neq i$.
- That is

$$\phi^{i,\alpha;j,\beta}(t) = \begin{cases} \phi^{i,\alpha;\beta}\ \ (t) \text{ if } i \neq j \\ \phi^{i,\alpha;i,\beta}(t) \text{ if } i = j \end{cases}$$

Data are labelled data provided by Euronext

- CAC40 index future
- we consider the most liquid expiry for each day
- from March 1st 2016 to February 28th 2017;
- 111 unique members (connections are aggregated);
- focus on equity hours (08:00 - 16:30 London time)

We consider this subset of agents:

- at least 1000 orders at the first level;
- are active “uniformly” between 08:00 and 16:30;
- respect the above for at least 30 days.

Total number of agent considered $M = 16$ **(**$+1$**)**

Name	Description
End of day (EOD) position	Absolute change in inventory at the end of the trading day, divided by the total volume traded by the agent.
Proprietary	Fraction of the orders that are market as proprietary by the agent.
Order lifetime	Median time between limit order insertion and cancellation/modification.
Inter-event time	Median time between two different orders by the same agent.
Limit-filled	Fraction of the submitted limit orders that are at least partially filled.
Canceled orders	Fraction of limit orders that are eventually canceled.
Aggressive volume	Ratio of the volume traded aggressively over the total traded volume by the agent.
Orders/Trades	Number of orders submitted for each trade.
Order size	Average order size (in contracts).
Time present at L1	Fraction of time the agent was present with a limit at at least one of the best quotes.
Present at both sides	Given the agent was present at the best, fraction of time he was present at both sides simultaneously.
Active connections	Average number of connections used by the agent per day.
Daily volume fraction	Fraction of the total traded volume (total buy + total sell) in which the agent is involved.

	240	140	478	127	636	398	503	274	566	59	584	364	597	455	244	669
EOD Position / Volume (%)	0.00	0.01	0.15	3.73	3.83	4.54	9.71	14.9	16.2	22.3	18.3	22.7	24.5	29.1	28.2	32.8
% Proprietary	100.0	100.0	100.0	100.0	100.0	100.0	0.22	68.1	97.8	100.0	1.19	100.0	2.10	98.7	0.00	0.37
Order lifetime (s)	0.51	0.61	0.20	3.57	0.99	0.33	1.33	3.19	42.0	4.14	7.87	5.17	4.32	3.04	6.31	11.1
Inter-event time (s)	0.01	0.00	0.02	0.01	0.00	0.06	0.63	0.07	0.63	0.01	1.64	0.01	1.65	0.12	2.45	2.33
Limit filled (%)	5.09	6.15	8.40	10.5	6.35	10.5	28.3	19.8	47.9	1.58	50.4	5.45	42.4	4.05	23.5	42.0
Limit (%)	51.1	50.0	48.9	44.3	36.3	37.7	49.3	47.5	53.6	40.7	31.0	51.0	54.1	50.0	48.1	53.4
Cancel (%)	48.4	47.2	46.2	40.0	33.7	33.9	36.2	37.9	27.9	40.1	14.5	48.3	31.1	48.0	36.6	30.2
Replace (%)	0.00	0.08	3.43	13.6	29.4	27.4	6.58	8.78	7.54	18.4	40.1	0.04	5.77	1.57	11.1	8.57
Aggressive (%)	0.51	2.69	1.42	2.08	0.60	1.01	7.97	5.76	10.9	0.80	14.4	0.62	9.08	0.43	4.25	7.78
Aggressive volume (%)	14.9	64.0	34.0	34.4	15.0	13.2	49.9	46.9	37.4	56.2	44.4	25.0	34.8	27.0	27.0	28.8
Orders/Trades (%)	3994.2	1085.8	1351.5	1128.0	5573.1	1036.1	238.5	524.7	190.3	5276.6	191.9	2609.4	206.5	3915.7	162.6	497.4
Order size (contracts)	1.02	1.38	2.33	1.65	1.15	4.41	2.45	1.64	1.70	1.88	3.66	2.42	2.70	4.08	3.75	2.38
Time present at L1 (%)	76.8	99.4	51.1	87.6	73.7	26.5	39.3	38.4	22.7	30.4	19.7	36.1	25.1	22.2	27.0	42.6
Present at both sides (%)	39.1	69.1	9.21	36.9	25.9	0.69	4.71	5.07	1.59	1.61	0.99	1.32	1.75	0.69	1.91	5.87
Active connections	19.9	98.2	16.2	32.2	19.6	2.16	18.9	19.8	9.32	5.47	17.7	10.5	4.26	13.9	2.55	3.69
Daily volume fraction (%)	2.22	31.3	4.68	6.30	1.28	3.59	6.05	5.63	2.04	4.76	3.85	2.00	1.88	2.13	2.65	2.73

	240	140	478	127	636	398	503	274	566	59	584	364	597	455	244	669
EOD Position / Volume (%)	0.00	0.01	0.15	3.73	3.83	4.54	9.71	14.9	16.2	22.3	18.3	22.7	24.5	29.1	28.2	32.8
% Proprietary	100.0	100.0	100.0	100.0	100.0	100.0	0.22	68.1	97.8	100.0	1.19	100.0	2.10	98.7	0.00	0.37
Order lifetime (s)	0.51	0.61	0.20	3.57	0.99	0.33	1.33	3.19	42.0	4.14	7.87	5.17	4.32	3.04	6.31	11.1
Inter-event time (s)	0.01	0.00	0.02	0.01	0.00	0.06	0.63	0.07	0.63	0.01	1.64	0.01	1.65	0.12	2.45	2.33
Limit filled (%)	5.09	6.15	8.40	10.5	6.35	10.5	28.3	19.8	47.9	1.58	50.4	5.45	42.4	4.05	23.5	42.0
Limit (%)	51.1	50.0	48.9	44.3	36.3	37.7	49.3	47.5	53.6	40.7	31.0	51.0	54.1	50.0	48.1	53.4
Cancel (%)	48.4	47.2	46.2	40.0	33.7	33.9	36.2	37.9	27.9	40.1	14.5	48.3	31.1	48.0	36.6	30.2
Replace (%)	0.00	0.08	3.43	13.6	29.4	27.4	6.58	8.78	7.54	18.4	40.1	0.04	5.77	1.57	11.1	8.57
Aggressive (%)	0.51	2.69	1.42	2.08	0.60	1.01	7.97	5.76	10.9	0.80	14.4	0.62	9.08	0.43	4.25	7.78
Aggressive volume (%)	14.9	64.0	34.0	34.4	15.0	13.2	49.9	46.9	37.4	56.2	44.4	25.0	34.8	27.0	27.0	28.8
Orders/Trades (%)	3994.2	1085.8	1351.5	1128.0	5573.1	1036.1	238.5	524.7	190.3	5276.6	191.9	2609.4	206.5	3915.7	162.6	497.4
Order size (contracts)	1.02	1.38	2.33	1.65	1.15	4.41	2.45	1.64	1.70	1.88	3.66	2.42	2.70	4.08	3.75	2.38
Time present at L1 (%)	76.8	99.4	51.1	87.6	73.7	26.5	39.3	38.4	22.7	30.4	19.7	36.1	25.1	22.2	27.0	42.6
Present at both sides (%)	39.1	69.1	9.21	36.9	25.9	0.69	4.71	5.07	1.59	1.61	0.99	1.32	1.75	0.69	1.91	5.87
Active connections	19.9	98.2	16.2	32.2	19.6	2.16	18.9	19.8	9.32	5.47	17.7	10.5	4.26	13.9	2.55	3.69
Daily volume fraction (%)	2.22	31.3	4.68	6.30	1.28	3.59	6.05	5.63	2.04	4.76	3.85	2.00	1.88	2.13	2.65	2.73

At the far left : Flat position, fast, high order to trade ratio, proprietary, high presence.
$\simeq$ **Market maker**

	240	140	478	127	636	398	503	274	566	59	584	364	597	455	244	669
EOD Position / Volume (%)	0.00	0.01	0.15	3.73	3.83	4.54	9.71	14.9	16.2	22.3	18.3	22.7	24.5	29.1	28.2	32.8
% Proprietary	100.0	100.0	100.0	100.0	100.0	100.0	0.22	68.1	97.8	100.0	1.19	100.0	2.10	98.7	0.00	0.37
Order lifetime (s)	0.51	0.61	0.20	3.57	0.99	0.33	1.33	3.19	42.0	4.14	7.87	5.17	4.32	3.04	6.31	11.1
Inter-event time (s)	0.01	0.00	0.02	0.01	0.00	0.06	0.63	0.07	0.63	0.01	1.64	0.01	1.65	0.12	2.45	2.33
Limit filled (%)	5.09	6.15	8.40	10.5	6.35	10.5	28.3	19.8	47.9	1.58	50.4	5.45	42.4	4.05	23.5	42.0
Limit (%)	51.1	50.0	48.9	44.3	36.3	37.7	49.3	47.5	53.6	40.7	31.0	51.0	54.1	50.0	48.1	53.4
Cancel (%)	48.4	47.2	46.2	40.0	33.7	33.9	36.2	37.9	27.9	40.1	14.5	48.3	31.1	48.0	36.6	30.2
Replace (%)	0.00	0.08	3.43	13.6	29.4	27.4	6.58	8.78	7.54	18.4	40.1	0.04	5.77	1.57	11.1	8.57
Aggressive (%)	0.51	2.69	1.42	2.08	0.60	1.01	7.97	5.76	10.9	0.80	14.4	0.62	9.08	0.43	4.25	7.78
Aggressive volume (%)	14.9	64.0	34.0	34.4	15.0	13.2	49.9	46.9	37.4	56.2	44.4	25.0	34.8	27.0	27.0	28.8
Orders/Trades (%)	3994.2	1085.8	1351.5	1128.0	5573.1	1036.1	238.5	524.7	190.3	5276.6	191.9	2609.4	206.5	3915.7	162.6	497.4
Order size (contracts)	1.02	1.38	2.33	1.65	1.15	4.41	2.45	1.64	1.70	1.88	3.66	2.42	2.70	4.08	3.75	2.38
Time present at L1 (%)	76.8	99.4	51.1	87.6	73.7	26.5	39.3	38.4	22.7	30.4	19.7	36.1	25.1	22.2	27.0	42.6
Present at both sides (%)	39.1	69.1	9.21	36.9	25.9	0.69	4.71	5.07	1.59	1.61	0.99	1.32	1.75	0.69	1.91	5.87
Active connections	19.9	98.2	16.2	32.2	19.6	2.16	18.9	19.8	9.32	5.47	17.7	10.5	4.26	13.9	2.55	3.69
Daily volume fraction (%)	2.22	31.3	4.68	6.30	1.28	3.59	6.05	5.63	2.04	4.76	3.85	2.00	1.88	2.13	2.65	2.73

At the far right :Slower, directional, lower order/trade ratio.
$\simeq$ **Directional agent**

Average direct and indirect contribution to the total volatility of a single event of type $m = \{i, \alpha\}$:

$$\xi_{i,\alpha}^2 = \left(\sum_j \sum_\beta \delta_{j,\beta} R^{j,\beta;i,\alpha} \right)^2$$

where we assume that $\delta_{j,\beta} = 0$ if $\beta \notin \{P^+, P^-\}$.

And the total diffusive volatility writes

$$\sigma^2 = \sum_{i,\alpha} \Lambda^{i,\alpha} \xi_{i,\alpha}^2$$

We construct a control result to compare with.

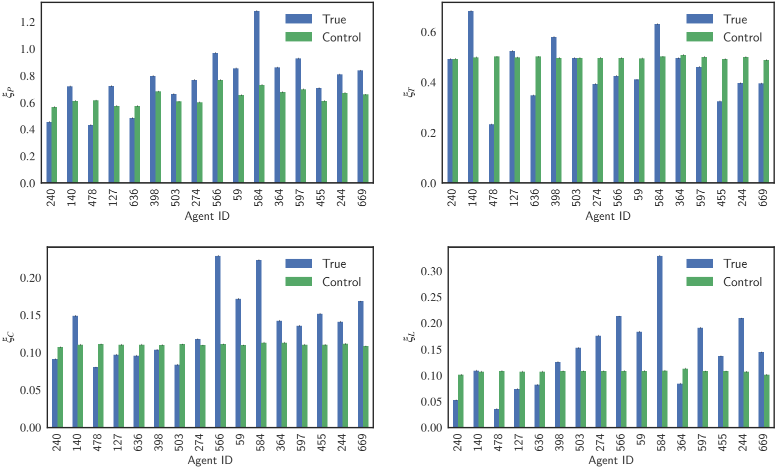
For each agent, each day the control has:

- Same number of orders.
- Same order composition.
- But agents labels are shuffled: orders are randomly assigned to agents.

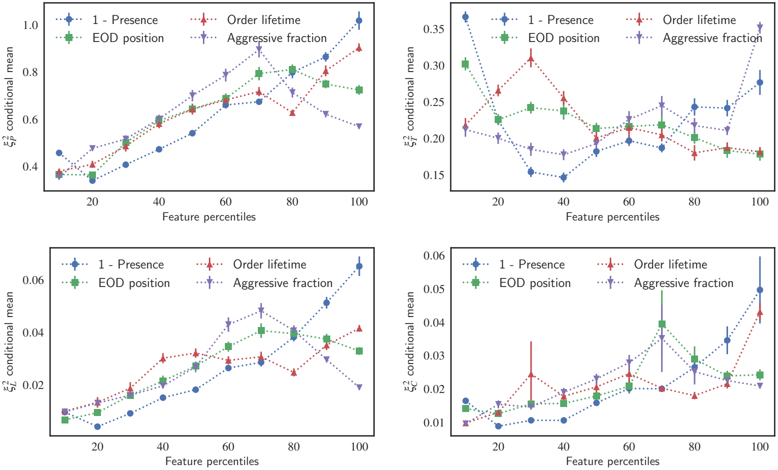
$\Longrightarrow$ **differences from control are mainly due to timing.**

Multi agent based Hawkes models

Volatility per event: agents averages



ξ essentially depends only on the agent order timing



Market maker like agents have smaller impact per passive event

Given

$$\sigma^2 = \sum_{i,\alpha} \Lambda^{i,\alpha} \left(\sum_{j,\beta} \delta_{j,\beta} R^{j,\beta;i,\alpha} \right)^2$$

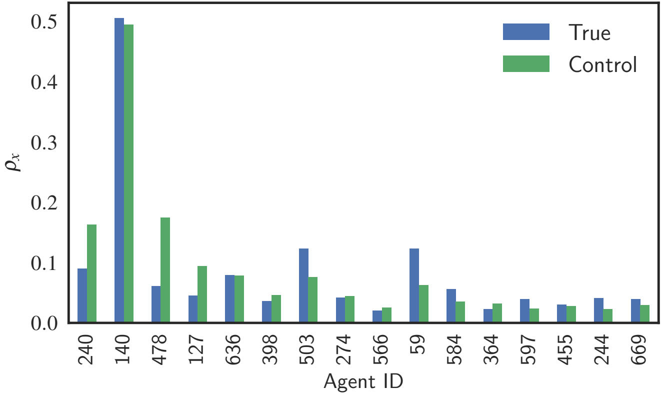
and

$$\Lambda^{i,\alpha} = \sum_{k,\gamma} R^{i,\alpha;k,\gamma} \mu^{k,\gamma},$$

we define ρ_m for agent m as

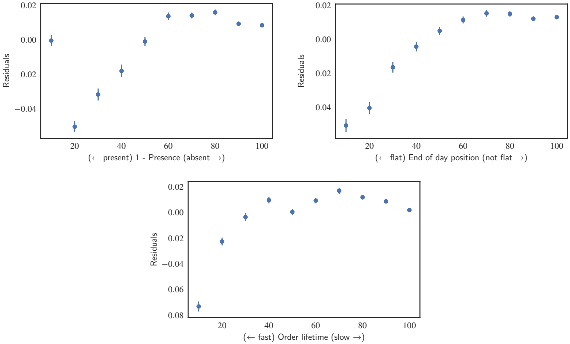
$$\sigma^2 \rho_m = \sigma^2 - \sum_{i \neq m} \sum_{k \neq m} \sum_{\alpha,\gamma} R^{i,\alpha;k,\gamma} \mu^{k,\gamma} \left(\sum_{j \neq m} \sum_{\beta} \delta_{j,\beta} R^{i,\alpha;j,\beta} \right)^2$$

ρ_m **: Relative difference in volatility we would observe if we removed all the activity directly or indirectly generated by agent** x**.**



Significant differences with the control for most agents (and $\rho_m > 0$)

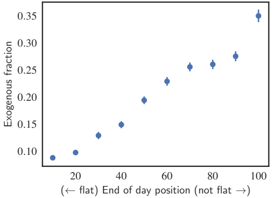
Plotting the residuals : $\rho_m - \rho_m^{\text{control}}$

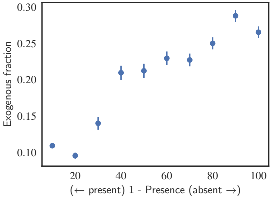


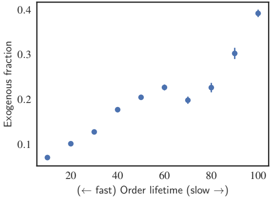
Market-marker like agent (left side) have volatility-attenuating behavior.

Exogenous fraction f_m for agent m

$$f_m = \frac{\sum_\alpha \mu^{m,\alpha}}{\sum_\alpha \Lambda^{m,\alpha}}.$$







Marker maker like (left side) are more "endogenous".

Towards the same approach for covariance (Wu, Bacry, Muzy, 2022)

- We now consider two stocks of the CAC40.
- Per agent, we use the exact same previous model "times" 2
- Thus for each agent, we get 8 dimensions (= # of order types) times 2 (for each stock)

$\Longrightarrow$

- "Crossing" between stocks are rare events
- Exogeneous covariance (due to covariations of the μ's) are most of the time small compared to Endogeneous covariance (due to agents behaviour)

(Ruan, Bacry, Deschatre, Hoffmann, Muzy, in preparation)

What if covariance comes from exogeneous unobserved information ?

- **Case 1** : There are some news impacting both stocks (but, I don't have access to them, and they happen at random times)
  $\Longrightarrow$ **Covariance induced by latent information**
- **Case 2** : Same agent is playing on two stocks, but we do not have access to labelled data
  $\Longrightarrow$ **Covariance induced by latent behavior**

$\Longrightarrow$ **We use a "shot noise" (a latent poisson process) that will influence the Hawkes process**

- the latent Poisson process is used to model some information arriving at random times (e.g., news)
- We keep the Hawkes process framework (with a latent component)

Event space $\mathcal{E} = \{P_1, P_2, X\}$ where

- processes N^{P_1} and N^{P_2} are observable (change of prices of stocks P_1 and P_2)
- process N^X is latent (e.g., the news)

The model is a 3d Hawkes model

$$\lambda^i(t) = \mu^i + \sum_{j \in \mathcal{E}} \phi^{ij}(t-s) dN^j \text{ for } i \in \{P_1, P_2\}$$

$$\lambda^k(t) = \mu^k_X \text{ for } k \in \{X\}$$

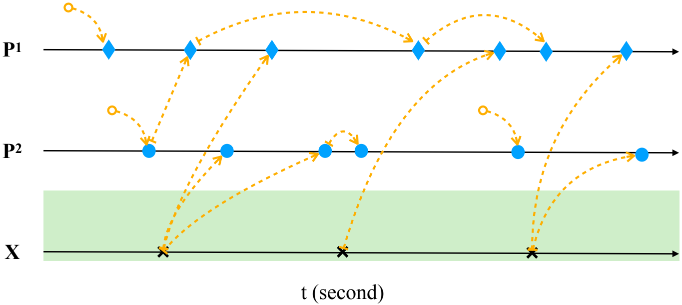


Figure: Illustration (2D : P_1, P_2)

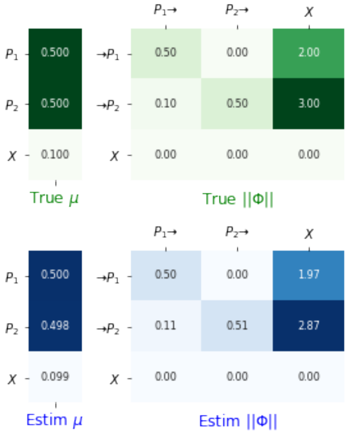


Figure: Estimation performed on a single realization 10^6s long ($\rho = 0.5$)

Correlation = 41%

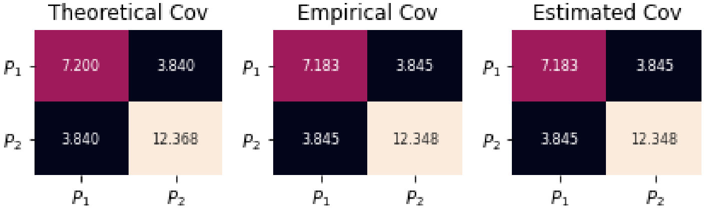
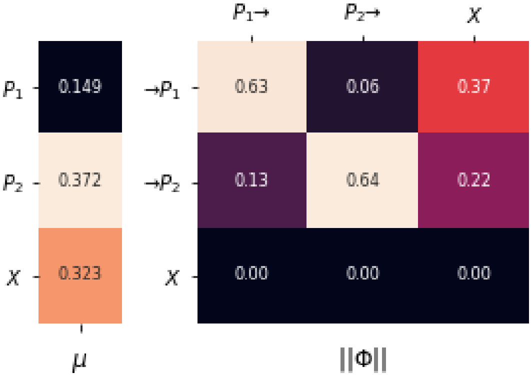
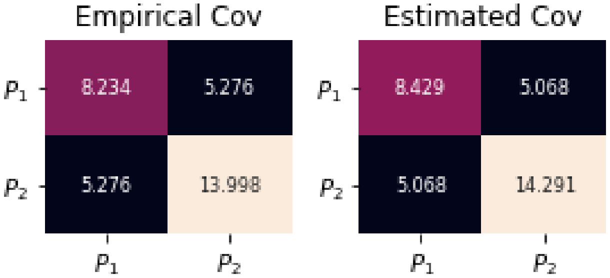


Figure: Covariance matrices


P_1
P_2
X
0.149
0.372
0.323
μ
$P_1\rightarrow$
$P_2\rightarrow$
X
$\rightarrow P_1$
$\rightarrow P_2$
X
0.63
0.06
0.37
0.13
0.64
0.22
0.00
0.00
0.00
$||\Phi||$

Correlation = 50%



- the latent Poisson models some explicit correlation between arrivals on P_1 and P_2
- We keep a 2d Hawkes process $\{N^{P_1}, N^{P_2}\}$
- Each time the latent component X jumps we add jumps to both components N^{P_1}, N^{P_2}, with some random delays

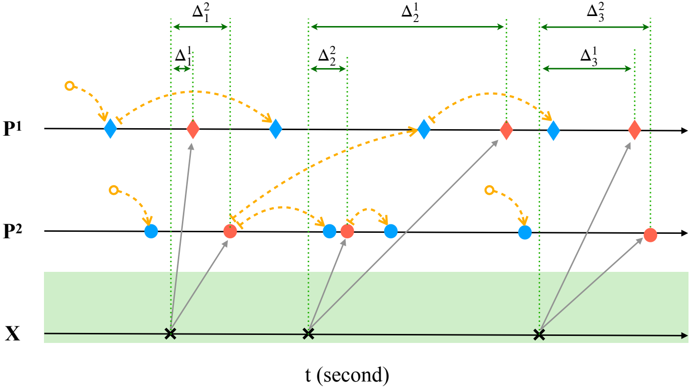
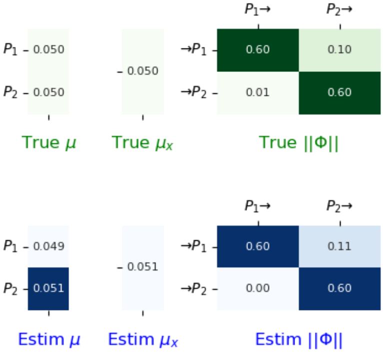


Figure: Illustration



Estimation on a single realization (10^6s, $\rho = 0.63$) - **Correlation 40%**

Covariance induced by latent behavior : Estimation

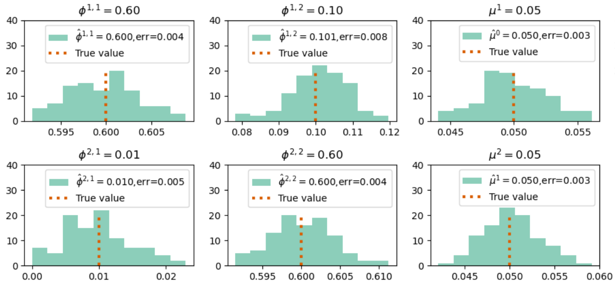
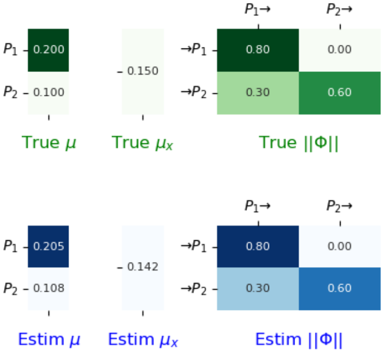


Figure: Parameters estimation distribution on 100 estimations



Estimation on a single realization (10^6s, $\rho = 0.63$) - **Correlation 80%**

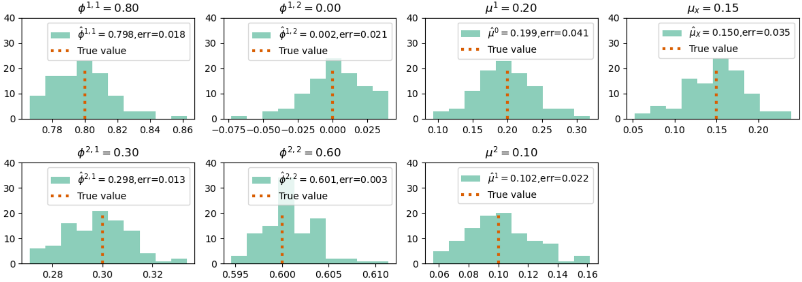
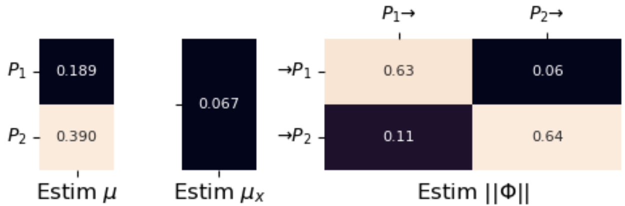


Figure: Parameters estimation distribution on 100 estimations



Correlation = 50%

M.Bompaire, P.Deegan, S.Gaiffas, S.Poulsen, E.B., ...

- Python 3 et C++11
- Open-source (BSD-3 License)
- `pip install tick` (on MacOS and Linux...)
- `https://x-datainitiative.github.io/tick`
- Statistical learning for time-dependent models
- Point processes (Poisson, Hawkes), Survival analysis, GLMs (parallelized, sparse, etc.)
- A strong simulation and optimization toolbox
- Partnership with Intel (use-case for new processors with 256 cores)
- Many contributors
- New contributors are welcome !

tick 0.1 Home Examples API Browse ▾ Search

tick

tick a machine learning library for Python 3. The focus is on statistical learning for time dependent systems, such as point processes. Tick features also tools for generalized linear models, and a generic optimization toolbox.

The core of the library is an optimization module providing model computational classes, solvers and proximal operators for regularization. It comes also with inference and simulation tools intended for end-users.

Show me »

Examples

Examples of how to simulate models, use the optimization toolbox, or use user-friendly inference tools.

Inference

User-friendly classes for inference of models

Simulation

User-friendly classes for simulation of data

Optimization

The core module of the library: an optimization toolbox consisting of models, solvers and prox (penalization) classes. Almost all of them can be combined