High frequency data
High frequency modeling and analysis

MASEF
Slides Part II
Tick by tick financial time series

Emmanuel Bacry
DR CNRS, CEREMADE, Université Paris-Dauphine, PSL,
CSO, Health Data Hub
bacry@ceremade.dauphine.fr

Tick by tick time series

- Non uniformly sampled (1d) time-series
- Which "tick" to choose ?
  - Last traded price
  - Mid price
  - Best bid/ask prices
  - ...

It is an arbitrary projection of a complex dynamics.

ACD Autoregressive Conditional Duration model (Engle, Russel 1997)

- Forex rate Mark/Dollar : 51 days May-August 1993
- best bid/ask prices - 303408 observations
- Average of 15s between each quote
- Strong intraday seasonality

ACD (Engle, Russel 1997) : intraday seasonality

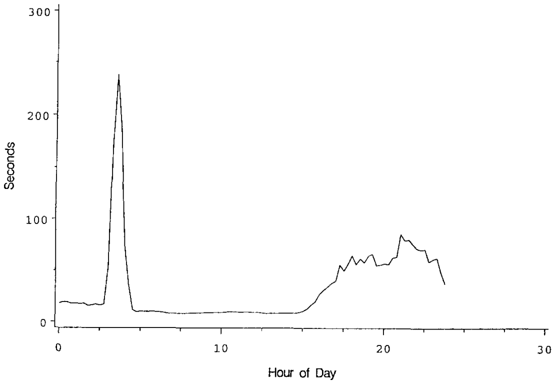


Fig. 1. Expected quote duration conditioned on time of day.

Robert F. Engle, Jeffrey R. Russell
Journal of Empirical Finance 4 (1997) 187-212.

$\Longrightarrow$ Simple "Deseasonalizing" by dividing the duration by the average duration

ACD (Engle, Russel 1997) : Autocorrelation function

	acf
lag 1	0.083
lag 2	0.076
lag 3	0.064
lag 4	0.053
lag 5	0.059
lag 6	0.048
lag 7	0.050
lag 8	0.038
lag 9	0.048
lag 10	0.040
lag 11	0.049
lag 12	0.043
lag 13	0.039
lag 14	0.040
lag 15	0.042

Robert F. Engle, Jeffrey R. Russell
Journal of Empirical Finance 4 (1997) 187-212.

ACD (Engle, Russel 1997) : The model

X_n : duration between two ticks

$$X_n = m_n \epsilon_n$$

where

- $\epsilon_n \geq 0$, iid
- $E(\epsilon_n) = 1$
- m_n independent from ϵ_n
- m_n $\mathcal{F}_{n-1}$ measurable

$$m_n = E(X_n | \mathcal{F}_{n-1}).$$

ACD (Engle, Russel 1997) : The model

$$m_n = E(X_n|\mathcal{F}_{n-1}).$$

$$m_n = \omega + \alpha X_{n-1} + \beta m_{n-1}$$

Thus

$$\begin{aligned} X_n &= m_n + X_n - m_n = \omega + \alpha X_{n-1} + \beta m_{n-1} + X_n - m_n \\ &= \omega + (\alpha + \beta) X_{n-1} - \beta (X_{n-1} - m_{n-1}) + X_n - m_n \end{aligned}$$

Asymptotically stationary if $\alpha + \beta < 1$

$$\Longrightarrow E(X_n) = \frac{\omega}{1 - (\alpha + \beta)} = M$$

$$\Longrightarrow X_n - M = (\alpha + \beta)(X_{n-1} - M) + X_n - m_n - \beta (X_{n-1} - m_{n-1})$$

ACD (Engle, Russel 1997) : The model

$$X_n - M = (\alpha + \beta)(X_{n-1} - M) + X_n - m_n - \beta(X_{n-1} - m_{n-1})$$

We set

- $Y_n = X_n - M$
- $W_n = X_n - m_n$: "innovation" (decorrelated with Y_{n-1})

ARMA(1,1) equation

$$Y[n] = (\alpha + \beta) Y_{n-1} + W_n - \beta W_{n-1}$$

ACD based models

Many many extensions ...

- ACM-ACD model (Russel,Engle 2004)
- $ACD(m, q)$ model
- EACD model
- log-ACD
- Burr-ACD
- GACD
- GARCH-ACD
- ...

(Co)Variance estimation and microstructure

Key issue : (historical) variance/covariance estimation

- diffusion processes : better estimates at fine scales
  $\Rightarrow$ one should use high frequency data (tick data)
- **However** : microstructure
  - Price processes are point processes
  - Prices "live" on a *tick grid*
  - **Strong mean reversion at very small scales**
  - Some references

    In economics : Roll (1984) [Roll model], Glosten (1987), Glosten et Harris (1988), Harris (1990)
    In statistics and econometrics : Gloter and Jacod (2001), Ait-Sahalia, Myland et Zhang (2002-2006)

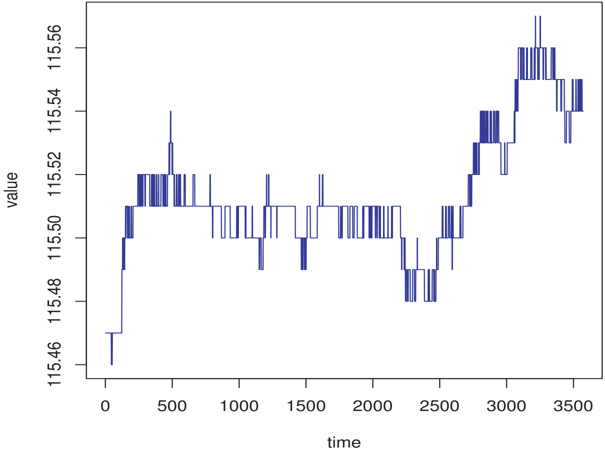


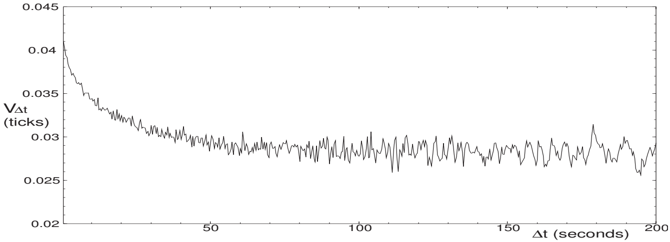
FIGURE: Bund 10Y, 6 Feb 2007, 09 :00–10 :00 (UTC) 1 data per second.

Variance estimators increase when going to high frequency

- $X(t)$: price (last traded price or mid-price or ...)
- Daily "variance" estimator :

$$V_{\Delta t} = \sum_{n=0}^{1day/\Delta t} |X((n+1)\Delta t) - X(n\Delta t)|^2$$

- Bund 10Y 21 days, 9-11 AM - Last Traded Ask - 7000 points



Hawkes processes : Mutually exciting point process models

- 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)
  $\longrightarrow$ Modeling high frequency time-series events (price changes, cancel/limit/market orders, ...)
- Rising popularity in machine learning (network, ...)

It's time to talk about Hawkes processes ....

A 2d Hawkes model for microstructure

E.B., S.Delattre, M.Hoffmann, J.F.Muzy
(Quant Finance 2012 + SPA 2013)

General form of the MEP price model

- $X_t = N_t^+ - N_t^-$ with

  $$N_t = \begin{pmatrix} N_t^+ \\ N_t^- \end{pmatrix}, \; \lambda_t^N = \begin{pmatrix} \lambda_t^{N^+} \\ \lambda_t^{N^-} \end{pmatrix}, \; u = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

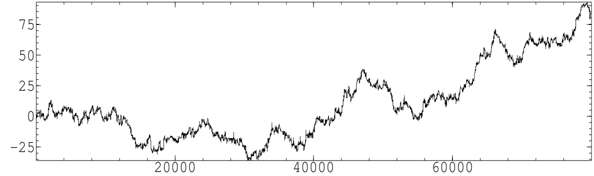
  $$\Phi^N(t) = \begin{pmatrix} \varphi^{\mathbf{N,s}}(\mathbf{t}) & \varphi^{N,c}(t) \\ \varphi^{N,c}(t) & \varphi^{\mathbf{N,s}}(\mathbf{t}) \end{pmatrix}$$

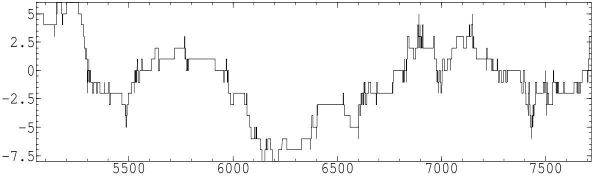
  $$\lambda_t^N = \mu.u + \Phi^N \star dN_t.$$

- Stability $\Longleftrightarrow \rho(||\Phi^N||) < 1$, where we defined

  $$||\Phi^N(t)|| = \begin{pmatrix} ||\varphi^{N,s}(t)||_1 & ||\varphi^{N,c}(t)||_1 \\ ||\varphi^{N,c}(t)||_1 & ||\varphi^{N,s}(t)||_1 \end{pmatrix}$$

Simulation over 10 hours + Zoom on 1h





Microstructure "Stylized facts"

$\longrightarrow$ Point processes (Hawkes) diffusing at large scales

$\longrightarrow$ Prices "live" on a *tick grid*

$\longrightarrow$ Strong mean reversion at very small scales

What about modeling two correlated assets ?

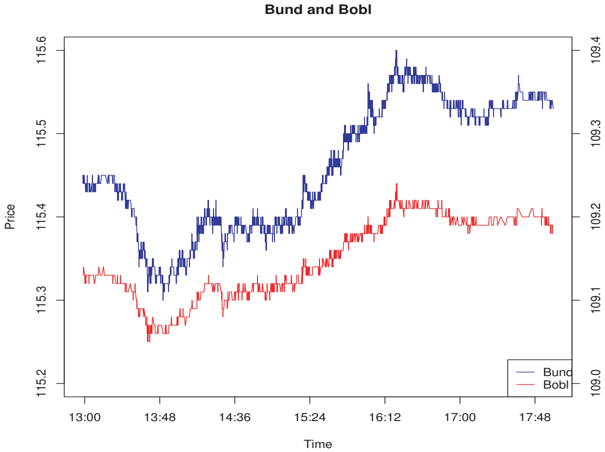
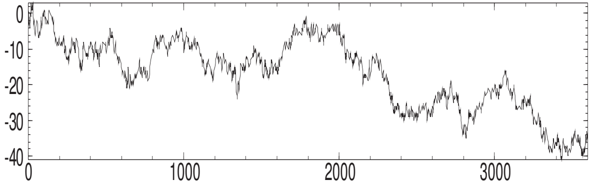
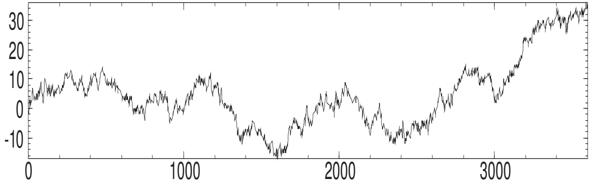


FIGURE: Bund 10Y / Bobl 5Y

A 4d Hawkes process for modeling 2 correlated assets

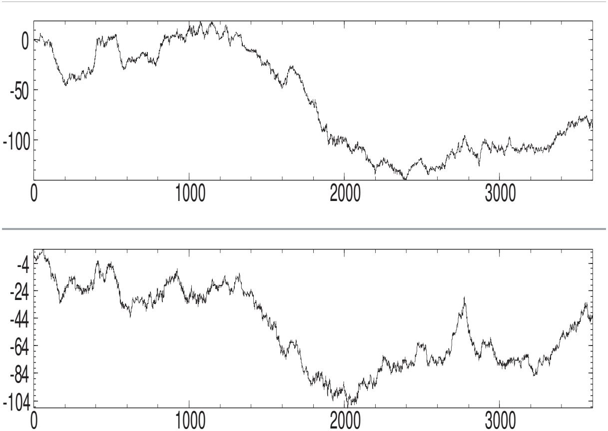
Diffusive correlation $C_{\Delta t=+\infty} = 10\%$





A 4d Hawkes process for modeling 2 correlated assets

Asymptotic correlation $C_{\Delta t=+\infty} = 60\%$



The mean signature plot in the dimension 1 model

- the signature plot :

$$V_{\Delta t} = \sum_{n=0}^{1day/\Delta t} |X((n+1)\Delta t) - X(n\Delta t)|^2$$

- the mean signature plot :

$$E(V_{\Delta t}) = \frac{1}{\Delta t} E(|X((n+1)\Delta t) - X(n\Delta t)|^2) = \frac{1}{\Delta t} E(X(\Delta t)^2)$$

with initial condition : $X(0) = 0$

- **closed-form formula for the mean signature plot** when $\Phi(x) = \alpha e^{-\beta x}$ (through the explicit computation of the Bartlett spectrum (1963)).

Closed form for the mean signature plot

Closed form for the mean signature plot

- $\lambda^{\pm}(t) := \frac{\mu}{2} + \int_{[0,t)} \phi(t-s) dN_s^{\mp}$
- $\phi(t) = \alpha e^{-\beta t} 1_{\mathbb{R}^+}(t), \quad ||\phi||_1 = \frac{\alpha}{\beta} < 1.$

$$E(V_{\Delta t}) = \Lambda \left[\nu^2 + (1-\nu^2) \frac{1 - e^{-\gamma \Delta t}}{\gamma \Delta t} \right],$$

where

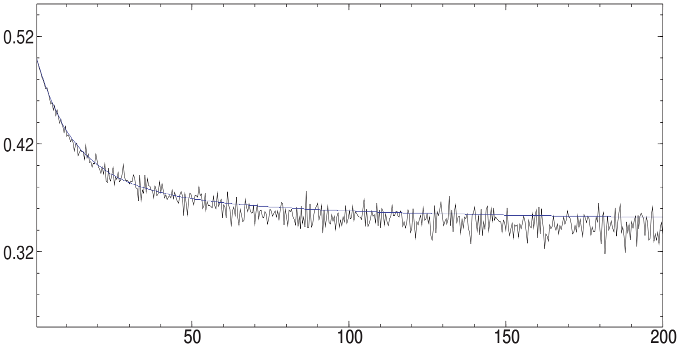
- $\Lambda = \frac{\mu}{1-||\phi||_1}$, $\nu = \frac{1}{1+||\phi||_1}$ and $\gamma = \alpha + \beta$

$\Longrightarrow$

- $E(V_{\Delta t = 0}) = \Lambda = 2E(\lambda^{\pm}) =$ "microstructural" variance
- $E(V_{\Delta t = +\infty}) = \Lambda \nu^2 =$ "diffusive" variance

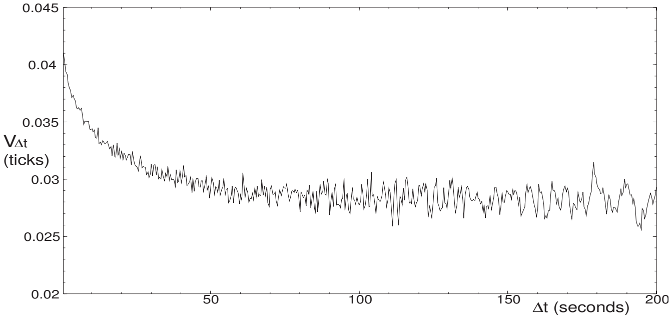
Mean signature plot on simulated data

Signature plot on 11 hours simulated data

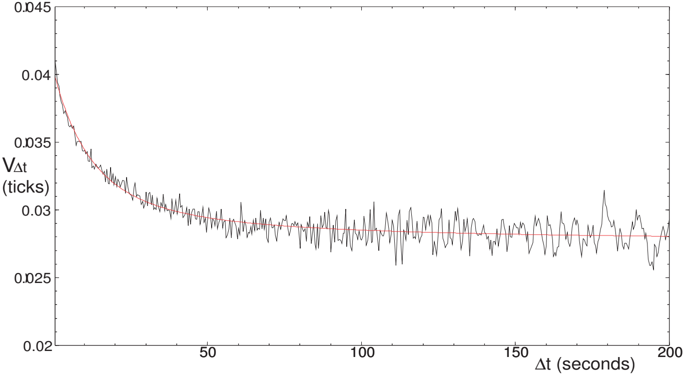


Mean signature plot on real data

- Bund 10Y : 21 days, 9-11 AM - Last Traded Ask (7000 points)



- Bund 10Y : 21 days, 9-11 AM - Last Traded Ask
  Mean square regression fit



$\Longrightarrow$ Very good modelization of the 1d microstructure noise.

The mean Epps effect the dimension 2 model

- Daily "correlation" estimator : $C_{\Delta t} = \tilde{C}_{\Delta t} / \tilde{C}_0$

$$\tilde{C}_{\Delta t} = \sum_{n=0}^{1day/\Delta t} (X((n+1)\Delta t) - X(n\Delta t))(Y((n+1)\Delta t) - Y(n\Delta t))$$

- the mean Epps effect

$$MEpps_{\Delta t} = \frac{E(X(\Delta t)Y(\Delta t))}{\sqrt{E(X(\Delta t)^2)E(Y(\Delta t)^2)}} \quad (1)$$

with initial condition : $X(0) = 0$

- **closed-form formula for the mean Epps effect** when $\Phi_{X,X}$, $\Phi_{Y,Y}$, $\Phi_{X,Y}$, $\Phi_{Y,X}$ are of the form $\alpha e^{-\beta x}$
$\rightarrow$ through the explicit computation of the **Bartlett spectrum (1963)**.

Closed form for the mean Epps effect in dimension 2

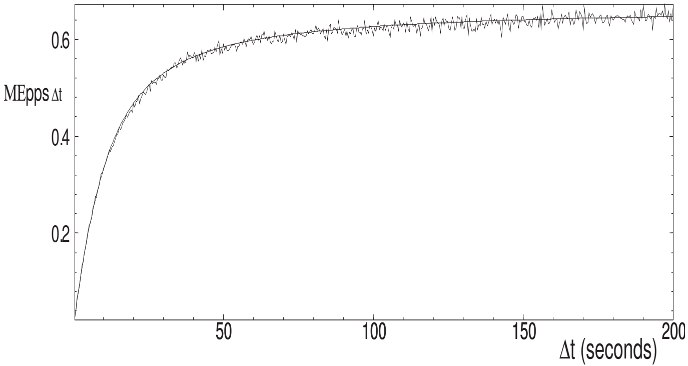
Closed form formula for the mean Epps effect in dimension 2

- General case $\rightarrow$ too many parameters
- Reducing the parameters
  - μ_X, μ_Y
  - $\alpha_{same} = \alpha_{X,X} = \alpha_{X,Y}$,
  - $\alpha_{cross} = \alpha_{X,Y} = \alpha_{Y,X}$,
  - $\beta = \beta_{X,Y} = \beta_{Y,X} = \beta_{X,X} = \beta_{Y,Y}$

$\longrightarrow$ Sorry : The formula is at least ... 6 slides long !

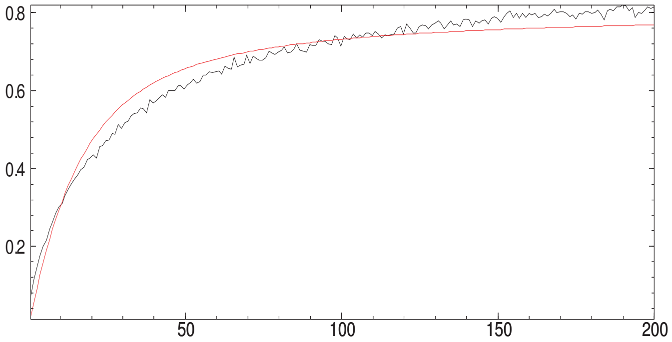
Mean Epps effect on simulated data

Mean Epps effect on 50 hours simulated data



Mean Epps effect on real data

- Bund 10Y / Bobl 5Y : 41 days, 9-11 AM - Last Traded



with $\alpha_{Bobl} = \alpha_{Bund}$.

Accounting for market impact of a labeled agent

- Agent at time t : $dA^+(t)$ buy orders $dA^-(t)$ sell orders

$$dA(t) = \begin{pmatrix} dA^+(t) \\ dA^-(t) \end{pmatrix}$$

- **Impacts will be modeled by additive terms on** λ^{N^+}, λ^{N^-}

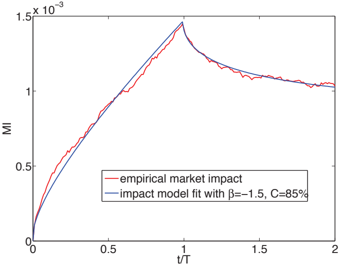
- **Single buy order** at time t_0 : $dA^+(t) = \delta(t - t_0)$, $dA^- = 0$
  - Impact on upward jumps : $\lambda_t^{N^+} \to \lambda_t^{N^+} + \varphi^{I,s}(t - t_0)$
    $\longrightarrow$ "Instantaneous" impact of the trade itself
    $\longrightarrow$ delayed upward moves (e.g., cancel orders)
  - Impact on donward jumps : $\lambda_t^{N^-} \to \lambda_t^{N^-} + \varphi^{I,c}(t - t_0)$
    $\longrightarrow$ delayed downward moves
- **Meta order** starting at time t_0

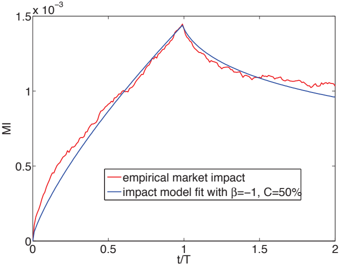
$$\lambda_t^N = \mu.u + \Phi^N \star dN_t + \Phi^I \star dA(t),$$

where $\Phi^I(t) = \begin{pmatrix} \varphi^{I,s} & \varphi^{I,c} \\ \varphi^{I,c} & \varphi^{I,s} \end{pmatrix}$ is the impact kernel

E.B, M.Hoffmann, A.Iuga, M.Lasnier, C.A.Lehalle (working paper)

$$MI(t) = E(X_t - X_{t_0=0}) \quad \text{with} \quad X_t = N_t^+ - N_t^-$$





- Concave impact while trading
  $\rightarrow$ **depend on T**, the smaller impact the larger T
- Relaxation after trading
- Is able to reproduce both permanent/non permanent impact

What if all available market orders are anonymous ?

Markets generally do not provide labeled data

- Flow of anonymous marker orders $T_t = \begin{pmatrix} T_t^+ \\ T_t^- \end{pmatrix}$

  T^+ (resp. T^-) : trade arrivals at the best ask (resp.bid)

- No more access to market impact profile !

- Only access to the Response function : $R(t - t_0)$:
  Expectation of the price at time t knowing there was a buying market order at time t_0, i.e.,

$$R(t - t_0) = E(N_t^+ - N_t^- \,|dT_{t_0}^+ = \delta(t - t_0))$$

Towards a model for market impact of anonymous market orders ?

Towards a model for market impact of anonymous market orders

Markets generally do not provide labeled data

- Flow of anonymous marker orders $T_t = \begin{pmatrix} T_t^+ \\ T_t^- \end{pmatrix}$

  T^+ (resp. T^-) : trade arrivals at the best ask (resp.bid)
- The Price model with a label agent

$$\lambda^N{}_t = \mu.u + \Phi^N \star dN_t + \Phi^I \star \mathbf{dA(t)}$$

- **The Price model with the anonymous market order flow**

$$\lambda^N{}_t = \Phi^N \star dN_t + \Phi^I \star \mathbf{dT(t)}$$

Towards a model for market impact of anonymous market orders

Markets generally do not provide labeled data

- Flow of anonymous marker orders $T_t = \begin{pmatrix} T_t^+ \\ T_t^- \end{pmatrix}$

  T^+ (resp. T^-) : trade arrivals at the best ask (resp.bid)
- The Price model with a label agent

$$\lambda^N{}_t = \mu.u + \Phi^N \star dN_t + \Phi^I \star \mathbf{dA}(\mathbf{t})$$

- **The Price model with the anonymous market order flow**

$$\lambda^N{}_t = \Phi^N \star dN_t + \Phi^I \star \mathbf{dT}(\mathbf{t})$$

E.B, J.F.Muzy (QF 2014)

- **The anonymous market orders flow** $\mathbf{T_t} = \begin{pmatrix} T_t^+ \\ T_t^- \end{pmatrix}$

  T^+ (resp. T^-) : trade arrivals at the best ask (resp.bid)

- **The Price model**

$$X_t = N_t^+ - N_t^-$$

  N^+ (resp. N^-) : upward (resp. downward) jumps

$$\lambda^N{}_t = \Phi^N \star dN_t + \Phi^I \star \mathbf{dT(t)}$$

$\longrightarrow$ Φ^I : "Instantaneous" impact + influence on price moves
$\longrightarrow$ Φ^N : Influence of past price moves on future price moves

The model for anonymous trades

E.B, J.F.Muzy (QF 2014)

The anonymous trade arrivals model $\longrightarrow$ A 2d Hawkes process

$$T_t = \begin{pmatrix} T_t^+ \\ T_t^- \end{pmatrix}$$

T^+ (resp. T^-) : trade arrivals at the best ask (resp.bid)

$$\lambda_t^T = \mu.u + \Phi^T \star dT_t + \Phi^R \star dN_t$$

where

$$u = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \; \Phi^T = \begin{pmatrix} \varphi^{T,s} & \varphi^{T,c} \\ \varphi^{T,c} & \varphi^{T,s} \end{pmatrix} \text{ and } \Phi^R = \begin{pmatrix} \varphi^{R,s} & \varphi^{R,c} \\ \varphi^{R,c} & \varphi^{R,s} \end{pmatrix}$$

$\longrightarrow$ μ : Anonymous trade intensity
$\longrightarrow$ Φ^T : Auto-correlation of trades
$\longrightarrow$ Φ^R : Retro-influence of price moves on trades

The overall model is a 4 dimensional Hawkes process P

E.B, J.F.Muzy (QF 2014)

$P_t = \begin{pmatrix} T_t \\ N_t \end{pmatrix}$ whose intensity $\lambda_t = \begin{pmatrix} {\lambda^T}_t \\ {\lambda^N}_t \end{pmatrix}$ is given by

$$\lambda_t = M + \Phi \star dP_t,$$

where

$$M = \begin{pmatrix} \mu.u \\ 0 \end{pmatrix}, \quad \Phi(t) = \begin{pmatrix} \Phi^T(t) & \Phi^R(t) \\ \Phi^I(t) & \Phi^N(t) \end{pmatrix}$$

- μ : Anonymous trade intensity
- $\Phi^T(t)$: Auto-correlation of anonymous trades
- $\Phi^I(t)$: "Instantaneous" impact + influence on price moves
- $\Phi^N(t)$: Influence of past price moves on future price moves
- $\Phi^R(t)$: Retro-influence of price moves on anonymous trades

Non parametric estimation

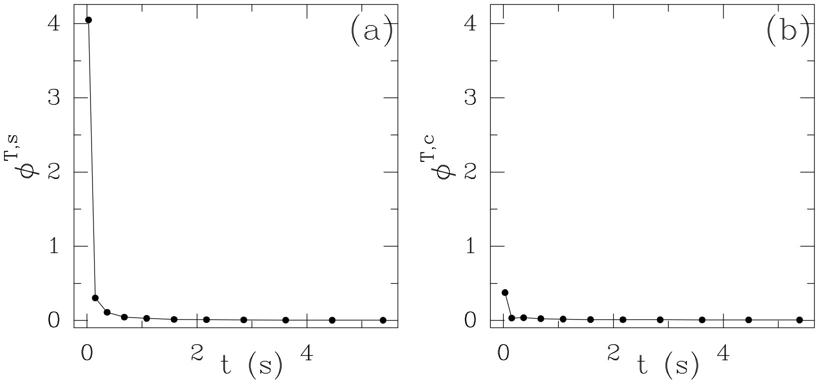
$$\lambda_t = M(t) + \Phi \star dP_t,$$

where

$$M(t) = \begin{pmatrix} \mu.u \\ 0 \end{pmatrix}, \quad \Phi(t) = \begin{pmatrix} \Phi^T(t) & \Phi^R(t) \\ \Phi^I(t) & \Phi^N(t) \end{pmatrix}$$

- **Non parametric estimation of** μ **and all the kernels** : Φ^T, Φ^R, Φ^N, Φ^I, **from anonymous market data**.

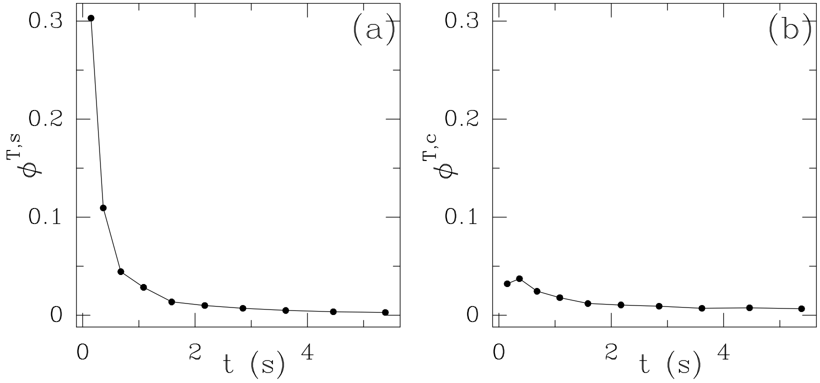
Non parametric estimation of Φ^T for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)



Trade auto-correlation

Non parametric estimation of Φ^T for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)

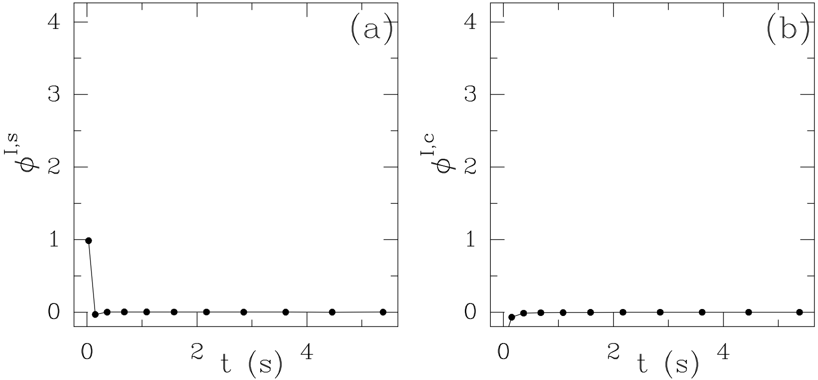
Zooming ...



Trade auto-correlation :

$\longrightarrow$ Mainly "positive" correlation : Splitting and Herding

Non parametric estimation of Φ^I for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)

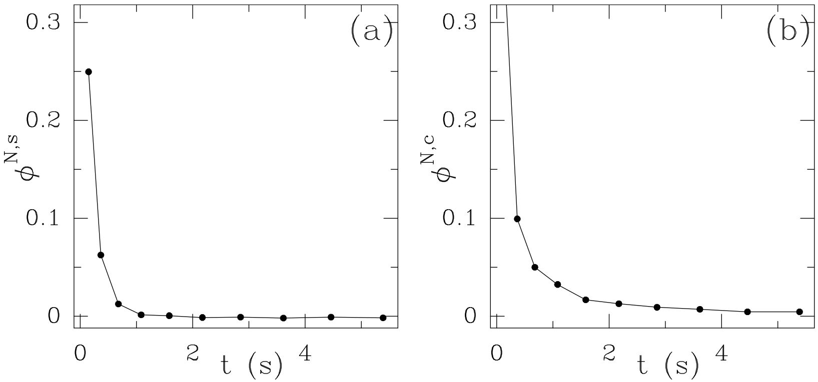


Trade "instantaneous" impact + influence on delayed price moves

$\longrightarrow$ Mainly instantaneous impact :

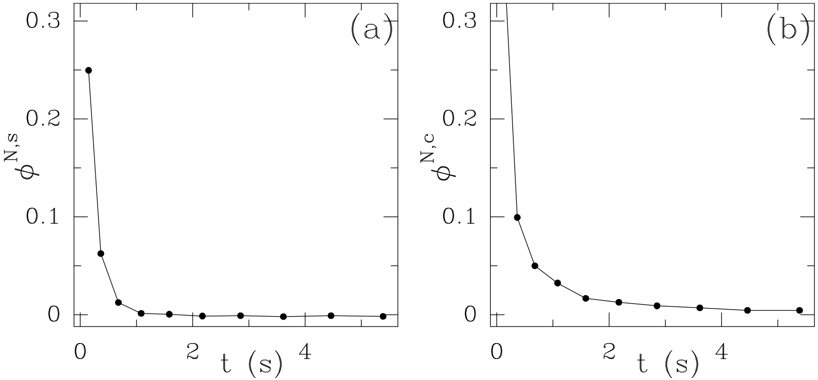
$\phi^{I,s}(t) \simeq C\delta(t)$ **and** $\phi^{I,c} \simeq 0$.

Non parametric estimation of Φ^N for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)



Influence of past price moves on future price moves

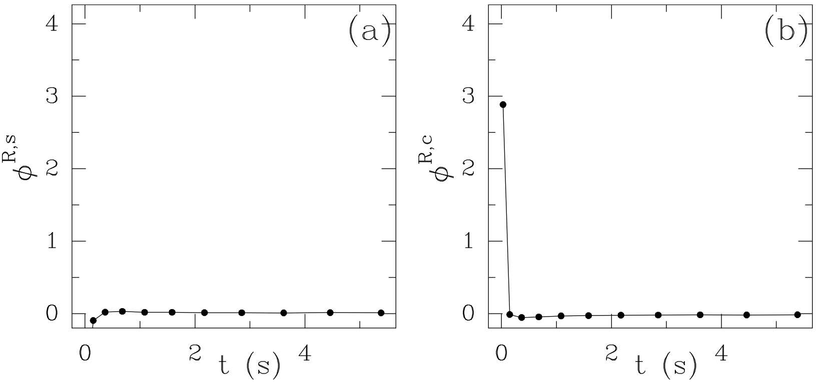
Non parametric estimation of Φ^N for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)



Influence of past price moves on future price moves

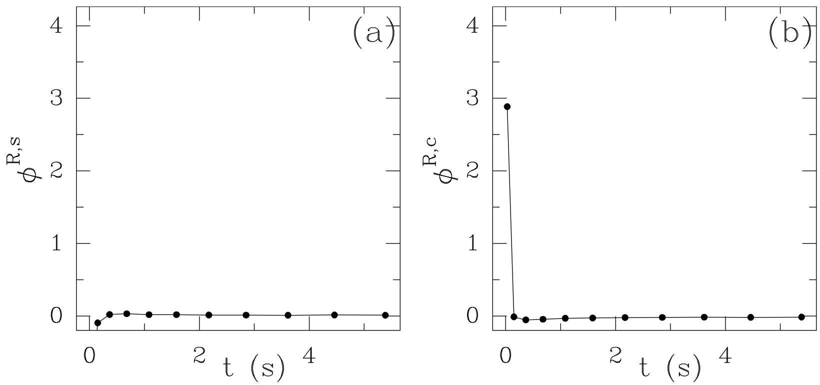
$\longrightarrow$ Mostly mean reverting

Non parametric estimation of Φ^R for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)



Retro-influence of price moves on anonymous trades :

Non parametric estimation of Φ^R for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)

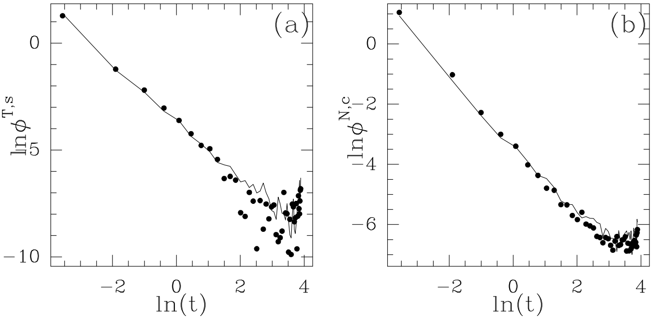


Retro-influence of price moves on anonymous trades :
$\longrightarrow \phi^{R,cross}$ large and $\phi^{R,self} \simeq 0$!

Price goes up $\Longrightarrow$ more sell market orders

Non parametric estimation for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)

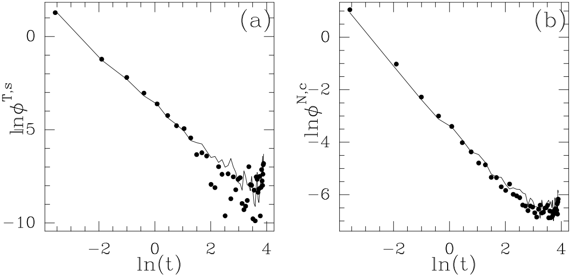
- **Most kernels are power-law** (when non 0) : $\frac{\alpha}{(\delta+x)^{\beta}}$
- **With** $\beta \simeq 1$: close to unstablity limit ! (K.Al Dayri, E.B, J.F.Muzy, EPJB 2012).



- Except $\varphi^{I,s} \simeq C\delta$ ($C << 1$) and $\varphi^{R,c}$

Non parametric estimation for Eurostoxx and Bund Futures 10h-12h, 2009-2012 (800 days)

- **Kernels can be amazingly stable when asset changes**

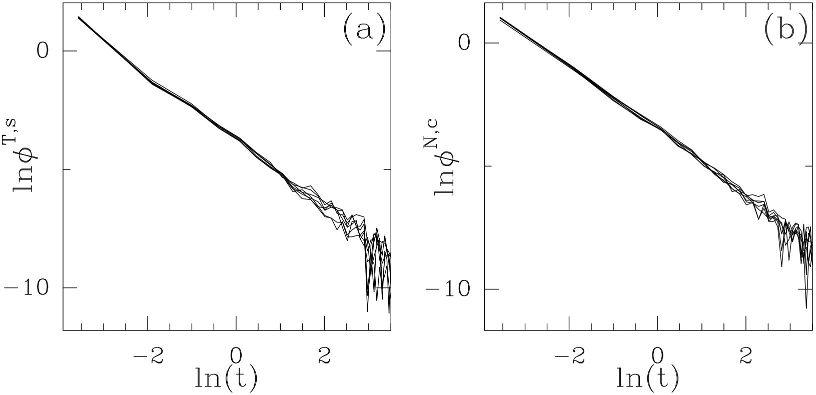


plain line : Eurostoxx Futures, • : Bund

- **No adjustment (no prefactors)** !

Intraday seasonalities for Eurostoxx Futures 2009-2012

Log-Log plots of $\varphi^{T,s}$ and $\varphi^{N,c}$ for different intraday slices :
9h-11h, 10h-12h, 11h-13h, 12h-14h, 13h-15h, 14h-16h, 15h-17h



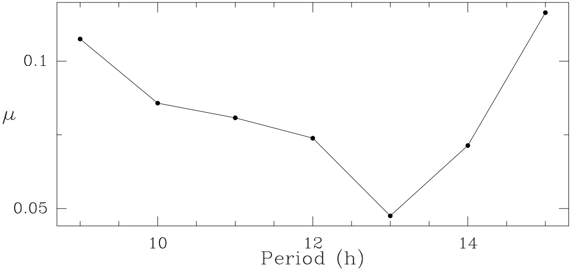
The kernel estimations do not depend on the intraday period

The intraday seasonality is only carried by μ (U-shape)

Model with intraday seasonality

$$\lambda_t = M(t) + \Phi \star dP_t,$$

where $M(t) = \begin{pmatrix} \mu_{\textbf{seasonal(t)}} \cdot u \\ 0 \end{pmatrix}, \quad \Phi(t) = \begin{pmatrix} \Phi^T(t) & \Phi^R(t) \\ \Phi^I(t) & \Phi^N(t) \end{pmatrix}$



Closed analytical formula for ....

Closed analytical formula for many quantities of interest

- Response function
- Diffusive variance of the price
- Auto-correlation function of
  - the trade signs (in practice heavy correlation)
  - the increments of the price (in practice very small correlations)
- Market impact
- ...

Analytical formula for the Market Impact of a meta-order

A particular case :

- An **"impulsive" Impact kernel** : $\varphi^{I,s}(t) = C\delta(t)$, $\varphi^{I,c}(t) = 0$
- A **single buy order** : $dA^+(t) = \delta(t), dA^-(t) = 0$

$\Longrightarrow$ the Market impact is

$$MI(t) = E(X_t - X_0) = 1_{[0,+\infty]}(t) - \int_0^t \Delta\xi(u)du,$$

where the Laplace transform of $\Delta\xi(t)$ is given by

$$\widehat{\Delta\xi} = 1 - \frac{(1 - \Delta\widehat{\phi}^T)}{(1 - \Delta\widehat{\phi}^T)(1 - \Delta\widehat{\phi}^N) - \Delta\widehat{\phi}^R}$$

where $\Delta\varphi^? = \varphi^{?,s} - \varphi^{?,c}$ measures the "kernel's imbalance"

Permanent versus non-permanent market impact

Analytical formula for the asymptotic market impact $MI(+\infty)$

In the case of a "cross-only" Retro-kernel : $\varphi^{R,s}(t) = 0$

$\Longrightarrow$ The asymptotic market impact is

$$MI(+\infty) = \frac{1}{(1 - \Delta||\varphi^N||_1) + ||\varphi^{R,c}||_1/(1 - \Delta||\varphi^T||_1)},$$

where

$$\Delta||\varphi^T||_1 = ||\varphi^{T,s}||_1 - ||\varphi^{T,c}||_1 \ (\in]-1,1[\text{ implied by stability})$$

$$\Delta||\varphi^N||_1 = ||\varphi^{N,s}||_1 - ||\varphi^{N,c}||_1 \ (\in]-1,1[\text{ implied by stability})$$

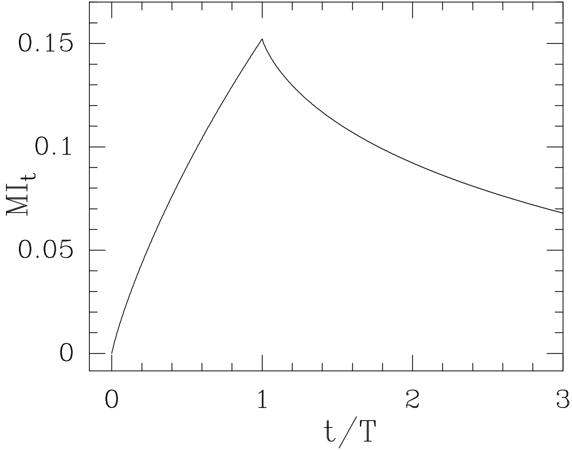
$MI(+\infty)$ **decreases when mean reversion increases, i.e. :**

- **when** $\Delta||\varphi^N||_1$ **goes to -1**
- **when** $||\varphi^{R,c}||_1$ **increases**
- **when** $\Delta||\varphi^T||_1$ **goes to 1**

Market impact profile estimation from anonymous data on Eurostoxx Futures

- Non parametric estimation of **all the kernels** : Φ^T, Φ^R, Φ^N, Φ^I
- Setting $\varphi^{T,c} = 0$, $\varphi^{I,c} = 0$ and $\varphi^{R,s} = 0$
- Fitting exponential kernels on $\varphi^{I,s}$ and $\varphi^{R,c}$
- Fitting Power-law kernels on $\varphi^{T,s}$, $\varphi^{N,c}$ and $\varphi^{N,s}$
- Computing the market impact profile from analytical formula

Market impact profile estimation from anonymous data on Eurostoxx Futures



Process diffusion

The process $P_t = \begin{pmatrix} T_t^- \\ T_t^+ \\ N_t^- \\ N_t^+ \end{pmatrix}$ **diffuses at large scales**

(from E.B., S.Delattre, M.Hoffmann, J.F.Muzy, preprint 2011)

$$\frac{1}{\sqrt{h}}(P_{ht} - E(P_{ht})) \rightarrow^{law} (\mathbb{I} - \widehat{\Phi}_0)^{-1}\Sigma^{1/2}W_t$$

where W_t is a n-dimensional Gaussian process (with stationary increments).

Consequently

- The *Trade process* $U_t = T_t^+ - T_t^-$ diffuses at large scales
- The *Price process* $X_t = T_t^+ - T_t^-$ diffuses at large scales

Trade sign long-range correlations

- U_t diffuses at large scales
- **Is it compatible with empirical findings about long range correlations of U_t ?**
  $\Rightarrow$ Strictly speaking : **NO** !

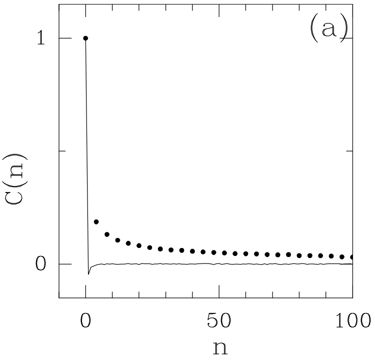
However, as long as

- $\Delta\widehat{\Phi^T}_0 \simeq 1$ and
- $\varphi^{T,s}{}_t \sim (c+t)^{-1+\nu}$,

$\Longrightarrow$ there is a finite range of scales (in practice $\simeq$ 5 decades !) on which

$$C^T(\tau) = Cov(U_t, U_{t+\tau}) \sim \tau^{2\nu-1}$$

Trade sign long-range correlations



- : $C^T(\tau) = Cov(U_t, U_{t+\tau})$

What about price efficiency ?

Price "long-memory puzzle" (Bouchaud etal. 2004) ?

- U_t is long-range correlated on a large range of scales
- How come the price $X_t = N_t^+ - N_t^-$ is not long-range correlated on a large range of scales ?

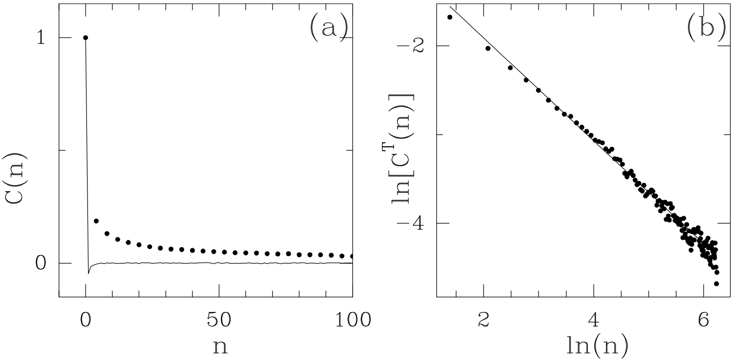
As long as

- $\Delta\widehat{\Phi^N}_0 < 0$ and
- $\Delta{\varphi^{N,s}}_t \sim (c' + t)^{-1+\nu'}, \quad (\nu' << 1)$,

$\Longrightarrow$ there is a finite range of scales (in practice $\simeq$ 5 decades !) on which

$$C^N(\tau) = Cov(X_t, X_{t+\tau}) << 1$$

Price fast decorrelation

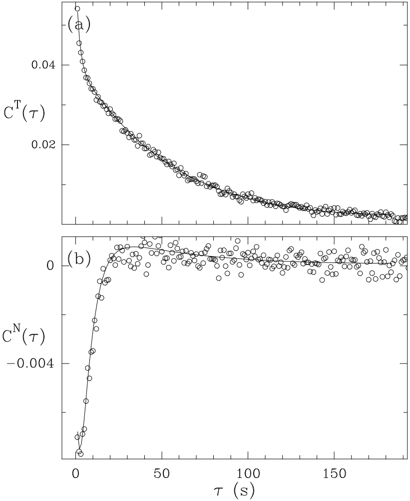


- (left and right plots) : $C^T(\tau) = Cov(U_t, U_{t+\tau})$

--- (left plot) : $C^N(\tau) = Cov(X_t, X_{t+\tau})$

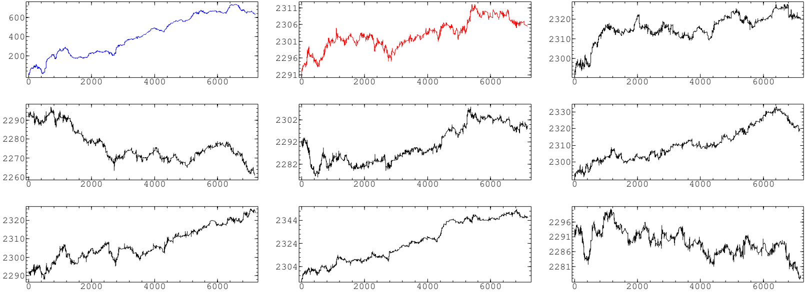
Trade sign and price correlations

(after removing the first point of correlation functions)



A microstructure and impact model

- Reproduce microstructure **and** market impact stylized facts (Bund, SP Fut., Euro/$ Fut., Eurostoxx Fut.,...)
- Kernel components can be easily estimated non parametrically
- **Most kernels are heavy-tailed** (as found in K.Al Dayri, E.B, J.F.Muzy, EPJB, 2012)
- **Kernel components can be easily interpreted in terms of various dynamics**
- **Analytical formula** for many quantities
- **Market impact profile estimation from anonymous data**
- Gives insights about the value of the permanent market impact
- Can be **easily generalized**
    - incorporating trade volumes
    - account for limit/cancel orders
    - Influence of labeled agents on anonymous agents
    - Multiple agents model
    - News model
- ...



$T_t^+ - T_t^-$: True cum. Trades on 3/08/2008 - [10am-12am]

$N_t^+ - N_t^-$: True mid-price on 3/08/2008 between 10am and 12am

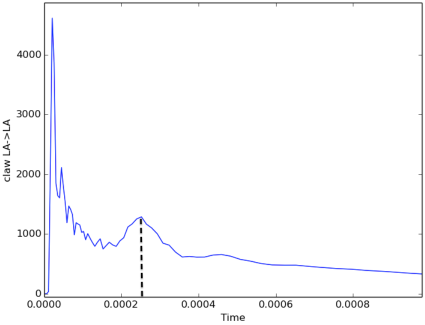
Simulation of the mid-price process N given the real trades

E.B. T.Jaisson and J.F. Muzy (2014)

- **Database** :
  - Dax Futures (small tick size)
  - Bund Futures (large tick size)
  - 1 year data : 06/2013-06/2014
  - **time precision = $\mathbf{1}\mu s$**
- P_t **is an 8-dimensional counting process** :
  - 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)

# events/day	PA/PB	TA/TB	LA/LB	CA/CB
Dax	72.000	20.000	152.000	184.000
Bund	14.000	28.000	240.000	212.000

Estimation is based on conditionnal expectation



$$E(dP_t^{LA} \mid dP_0^{LA} = 1)$$

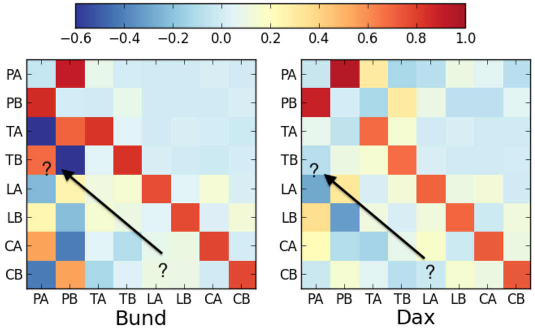
MOST of the conditionnal laws display a peak around $t \simeq 0.25ms$
$\Longrightarrow$ **Average Latency**

Level of exogeneity

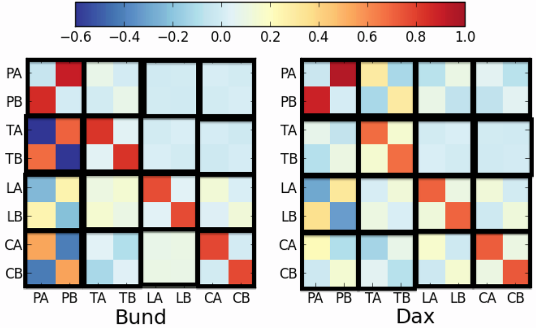
Ratio of exogeneous events over all events $R^i = \frac{\mu^i}{\Lambda^i}$

	PA	PB	TA	TB	LA	LB	CA	CB
Bund	4.4%	4.4%	4.5%	4.5%	1.4%	1.4%	1.6%	1.8%
Dax	2.7%	2.7%	4.3%	4.5%	1.1%	1.2%	0.7%	0.4%

Color coding of the norms $||\Phi^{?\to ?}||_1$

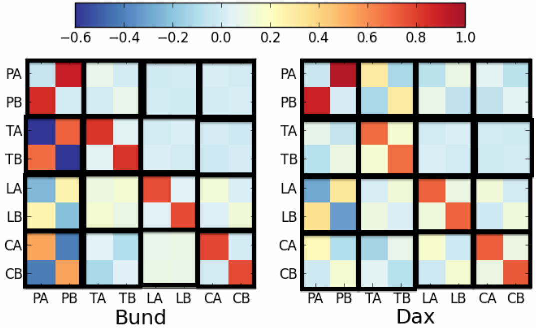


Kernel Norm



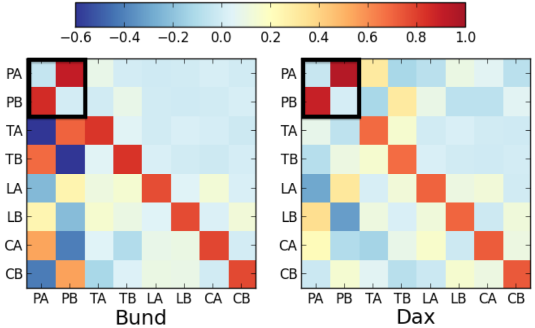
⟹ Symmetry upward/downward and ask/bid

Kernel Norm



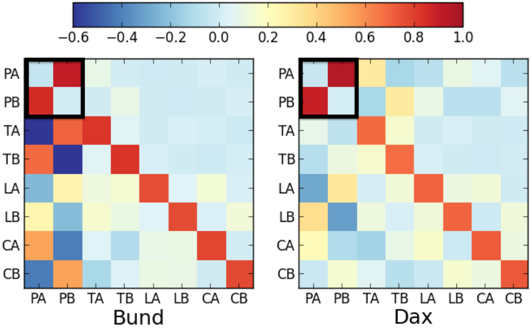
⟹ **Symmetry upward/downward and ask/bid**

Price Kernel Norms



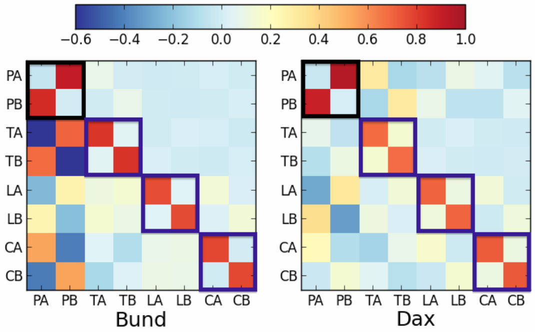
- "Anti-diagonal" shape in the price kernels
  $\Rightarrow$ **mean reversion of the price**

Price Kernel Norms



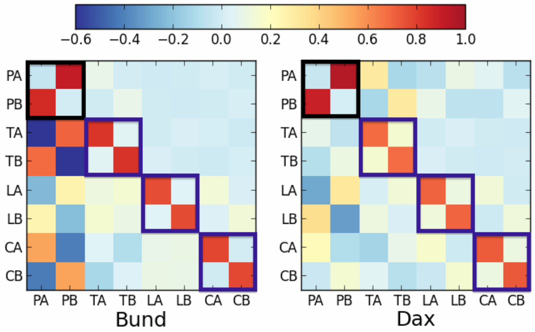
- "Anti-diagonal" shape in the price kernels
  $\Rightarrow$ **mean reversion of the price**

Order flow Kernel Norms



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

Order flow Kernel Norms

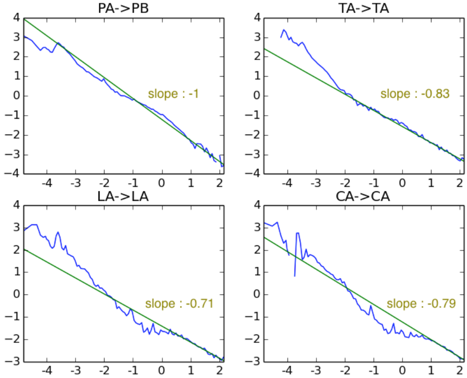


- "Anti-diagonal" shape in the price kernels
  $\Rightarrow$ **mean reversion of the price**
- "Diagonal" shape in the limit/cancel/trade kernels
  $\Rightarrow$ **splitting/herding**

Shape of some kernels

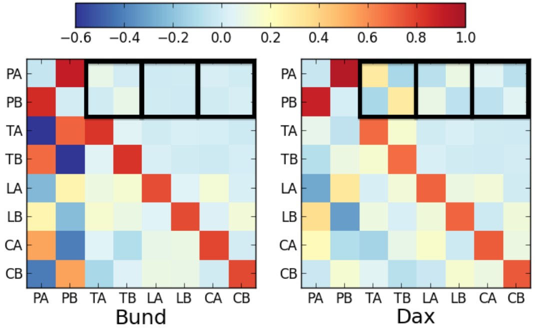
Power law kernels responsible for

- **price mean reversion**
- **order splitting, herding**



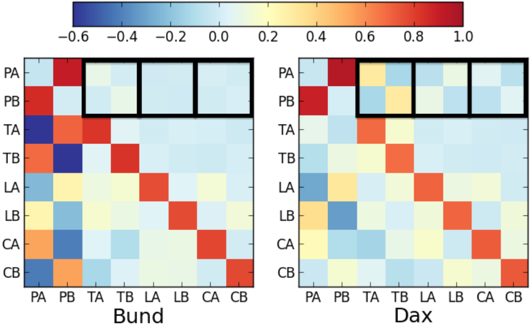
Log-log plots of some kernel estimations on 7 decades

Impact of the order flows on the price



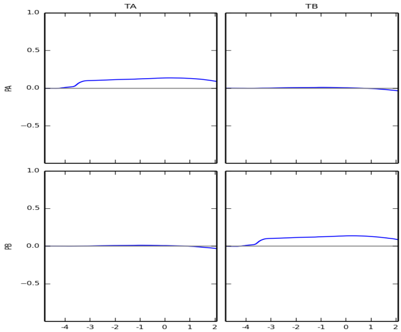
- Trades : main source of impact (diagonal)
- Limits : contrariant
- Cancels : diagonal

Impact of the order flows on the price



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

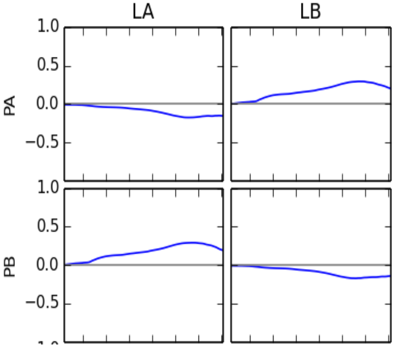
Price impact of trade flow



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" $\simeq 0.25ms$

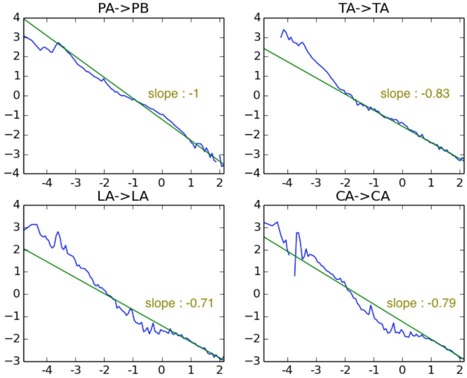
Price impact of limit/cancel flow



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

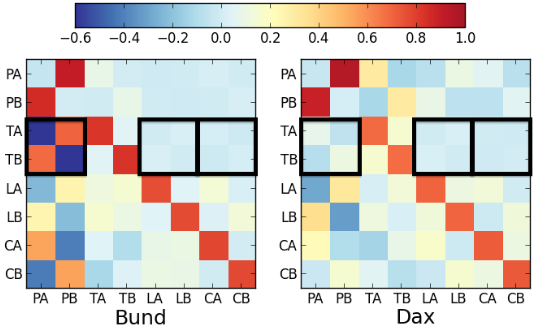
- The kernels $\Phi^{L?\to P?}$ and $\Phi^{C?\to P?} << \Phi^{T?\to P?}$
- The kernels $\Phi^{L?\to P?}$ and $\Phi^{C?\to P?}$ are **not** localized.

Market Price "efficiency"



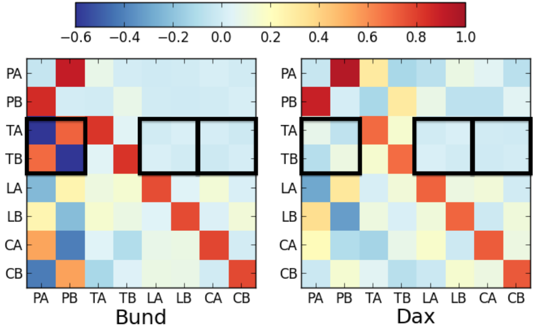
⇒ Market Price efficiency comes from a "rough" equilibrium between the 4 main power law kernels

Impact on the trades



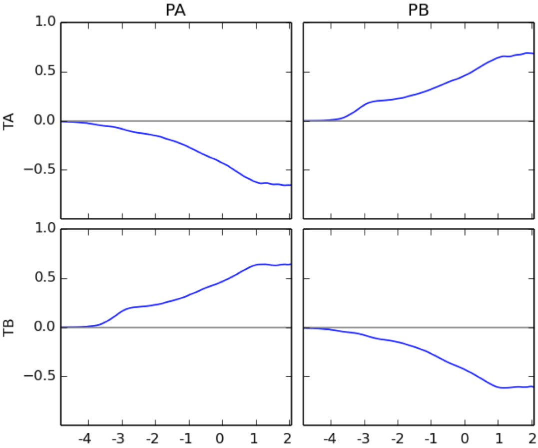
- Impact of the Price : Bund (contrariant), Dax (diagonal) large tick size : a change in price carries much more information
- Impact of the Limit is very small (actually trades are *leading*)
- Impact of the Cancel is very small

Impact on the trades



- Impact of the Price : Bund (contrariant), Dax (diagonal) large tick size : a change in price carries much more information
- Impact of the Limit is very small (actually trades are *leading*)
- Impact of the Cancel is very small

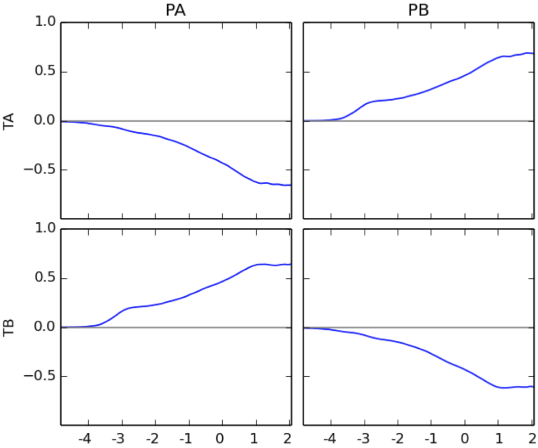
Impact of the price on trade flow of the Bund



Cumulative kernels $\int_0^t \Phi(s)ds$ as a function of $\log(t)$

- **Price goes up $\Rightarrow$ agents buy less and sell more**

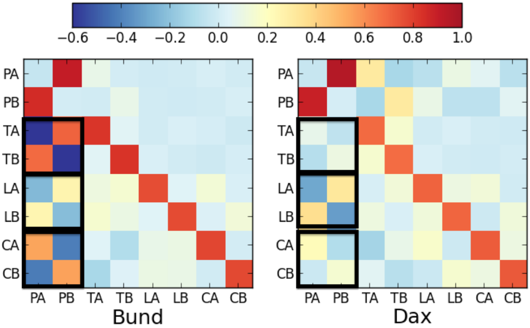
Impact of the price on trade flow of the Bund



Cumulative kernels $\int_0^t \Phi(s)ds$ as a function of $\log(t)$

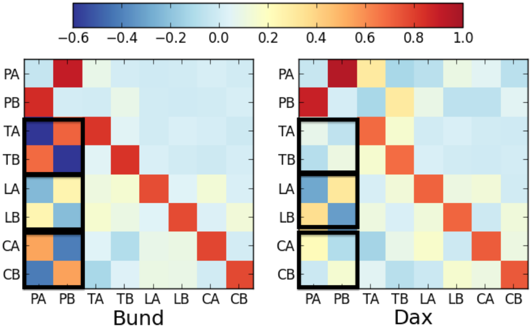
- **Price goes up $\Rightarrow$ agents buy less and sell more**

Impact of the price on order flows



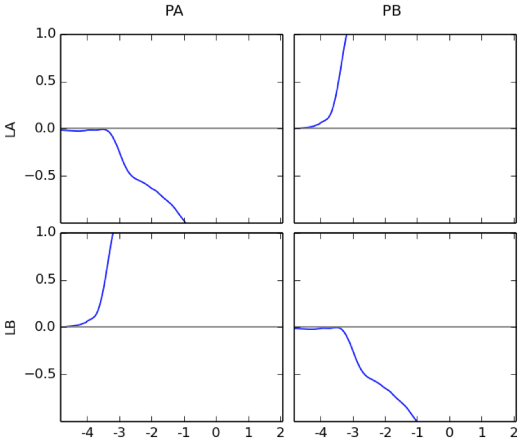
- Inpact on Trade : Bund (contrariant), Dax (diagonal)
- Impact on Limit : contrariant
- Impact on Cancel : diagonal

Impact of the price on order flows



- Inpact on Trade : Bund (contrariant), Dax (diagonal)
- Impact on Limit : contrariant
- Impact on Cancel : diagonal

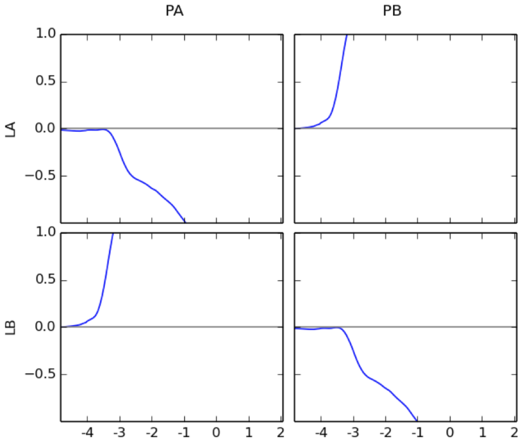
Impact of the price on limit flow of the Bund



Cumulative kernels $\int_0^t \Phi(s)ds$ as a function of $\log(t)$

- Price goes up $\Rightarrow$ Market maker reaction

Impact of the price on limit flow of the Bund



Cumulative kernels $\int_0^t \Phi(s)ds$ as a function of $\log(t)$

- **Price goes up $\Rightarrow$ Market maker reaction**

Conclusion and Prospects

- Kernel components can be easily estimated non parametrically
- Stable even for slightly negative valued kernels
- **Kernel components can be easily interpreted in terms of various dynamics**
  - Latency appears clearly in some kernels
  - Mean-reversion of price
  - Strong localized price impact of trades
  - Very weak non-localized price impact of limits and cancels
  - Contrariant impact of price changes on trade flow
  - Market maker reactions to price change
  - . . .