Hawkes Processes

Agenda

1. Why Hawkes Processes?
2. Point Processes
3. Self-Exciting Point Processes
4. Hawkes processes
   - Definition
   - Immigration-Birht Representation
   - Properties
5. Application - *high frequency price data*
   - Fitting high frequency realized volatility

Stock Price Models - A Hall of Fame

Geometric Brownian Motion

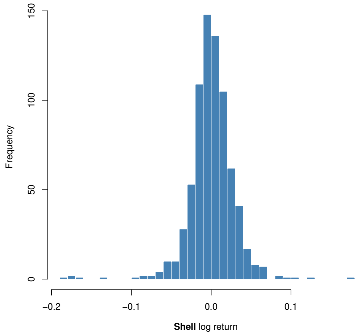
Let $\{S_t : t \geq 0\}$ defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with $\{\mathcal{F}_t : t \geq 0\}$ governed by,

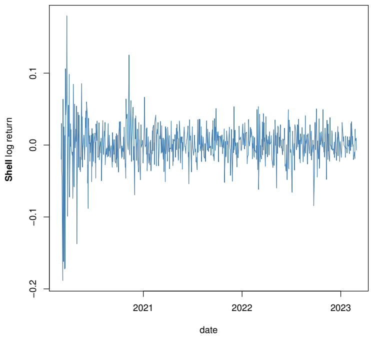
$$\frac{\mathrm{d}S_t}{S_t} = \mu \mathrm{d}t + \sigma \mathrm{d}W_t, \quad S_0 = x > 0.$$

Merton Jump-Diffusion Process

Let $\{S_t : t \geq 0\}$ defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with $\{\mathcal{F}_t : t \geq 0\}$ governed by,

$$\frac{\mathrm{d}S_t}{S_t} = (r - \lambda\bar{k})\mathrm{d}t + \sigma \mathrm{d}W_t + k\mathrm{d}N_t$$





Background

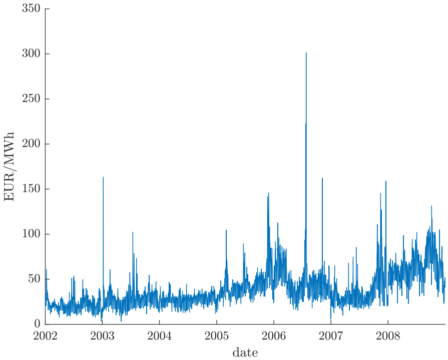


Figure: EPEX Spot Prices.

Our Spot Price Model

Let S_t denote the spot price,

$$S_t = \Lambda(t) + X_t + Y_t,$$

where $\{X_t : t > 0\}$ is an OU process defined by the SDE,

$$\mathrm{d}X_t = -\alpha X_t \mathrm{d}t + \sigma \mathrm{d}B_t,$$

$\{Y_t : t > 0\}$ is determined by,

$$\mathrm{d}Y_t = -\beta Y_t \mathrm{d}t + dL_t.$$

$\beta \in \mathbb{R}^+$ and L_t is a square integrable Lévy process.

Background

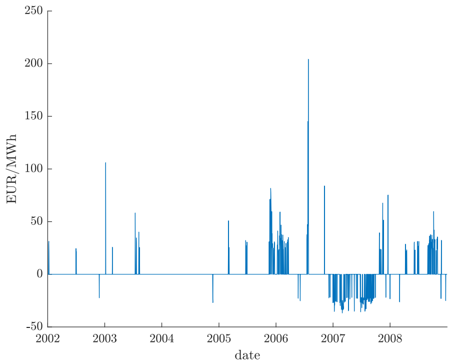


Figure: Filtered Spikes.

The Kou-Model

L_t is compound Poisson i.e.

$$L_t = \sum_{i=1}^{N_t} D_i,$$

where $N_t \sim \text{Pois}(\lambda)$ i.e. interarrival times of spikes are exponentially distributed. If $\{t_1, t_2, \ldots, t_N\}$, then $t_i - t_{i-1} \sim \text{Exp}(\lambda)$ for $i = 2, 3, \ldots, N$. The jump sizes D_i will be i.i.d. and double exponential distribution with density,

$$f_D(x) = p\eta_1 e^{-\eta_1 x} \mathbb{1}_{x \geq 0} + q\eta_2 e^{-\eta_2 |x|} \mathbb{1}_{x \leq 0},$$

with $p + q = 1$ and $\eta_1, \eta_2 \geq 0$

Background - Point Process

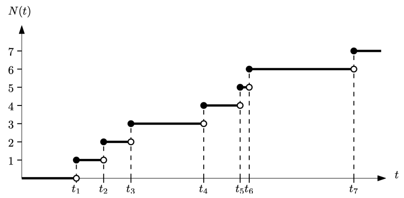


Figure: Point Process image taken from[a].

Poisson Process

- $t_1, t_2, \ldots, t_n$ i.i.d.
- $\lambda(t) = \lambda$ deterministic

Point Processes

On $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with $\{\mathcal{F}_t : t \geq 0\}$ then $N(t) \in \mathbb{Z}^+$ is a point process. Heuristically,

$$\begin{aligned}
\lambda(t) &= \lambda\left(t \mid \mathcal{F}_t\right) \\
&= \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}\left(\text{event takes place in } [t, t+h] \mid \mathcal{F}_t\right) \\
&= \lim_{h \downarrow 0} \frac{\mathbb{E}[N(t+h) - N(t) \mid \mathcal{H}(t)]}{h} \\
&= \text{"}\mathbb{E}\left[dN_t \mid \mathcal{F}_t\right]\text{"}
\end{aligned}$$

[a] Patrick J. Laub, Thomas Taimre, and Philip K. Pollett. *Hawkes Processes*. 2015.

Background - Point Process

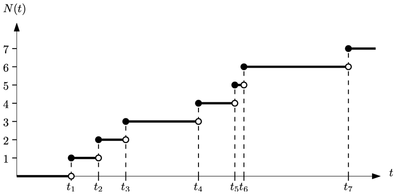


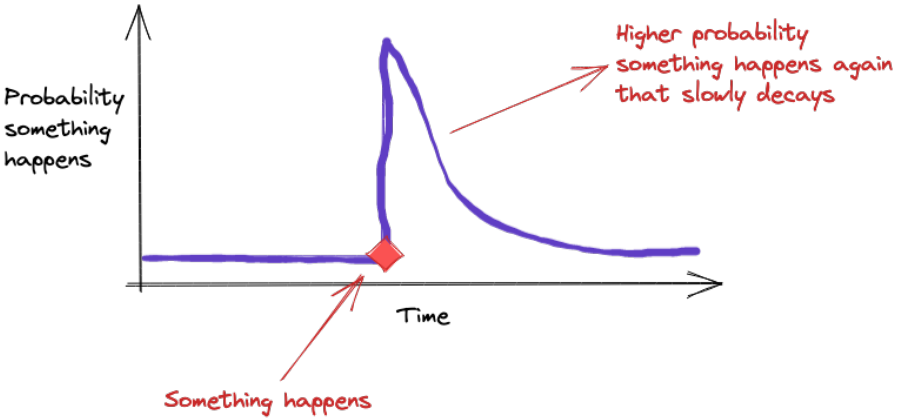
Figure: Point Process image taken from[a].

[a] Laub, Taimre, and Pollett, *Hawkes Processes*.

Properties of a Poisson Process

- Independent arrivals, $t_1, t_2, \ldots, t_n$ i.i.d.
- $\lambda(t) = \lambda$ deterministic
- $\mathbb{P}\left(N(T) = k|T\right) = \frac{(\lambda T)^k \exp -\lambda t}{k!}$
- $\mathbb{E}\left[N(T)\right] = \lambda T$
- $\mathbb{E}\left[N(T)^2\right] - \mathbb{E}\left[N(T)\right]^2 = \lambda T$
- Clustering Ratio
  - $\rho = \frac{\mathbb{E}\left[N(T)^2\right] - \mathbb{E}\left[N(T)\right]^2}{\mathbb{E}\left[N(T)\right]} = 1$
    - $\rho < 1$ - Repulsive Events, anti-correlated (negatively correlated)
    - $\rho > 1$ - bursts of activity, correlated as they have ρ larger than for a Poisson

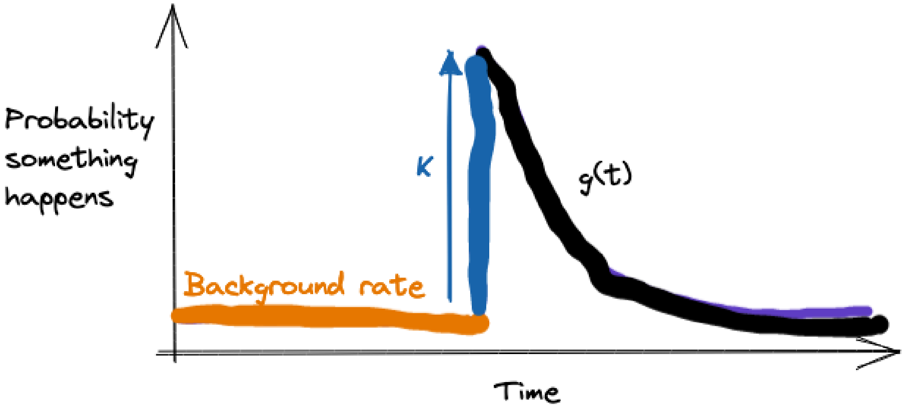
Self-Exciting Point Process[1]



[1] Dean Markwick. *Modelling Microstructure Noise Using Hawkes Processes*. URL: https://dm13450.github.io/2022/05/11/modelling-microstructure-noise-using-hawkes-processes.html.

Self-Exciting Point Process[2]

- $\lambda(t) = \mu + \kappa \sum_{t_i < t} g\left(t - t_i\right)$



[2] Markwick, *Modelling Microstructure Noise Using Hawkes Processes*.

Definition and Conditional Intensity

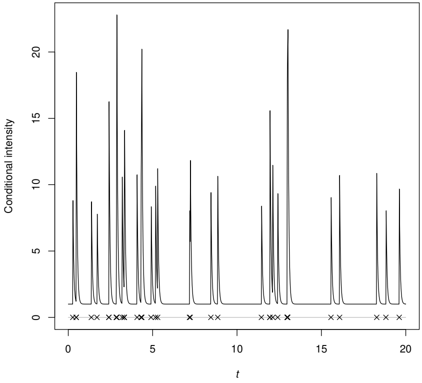


Figure: $\lambda^*(t) = 1 + 0.5 \cdot 20 \exp(-20t) 1_{\{t \geq 0\}}$.

Definition of a Hawkess Process

Suppose the process' conditional intensity function is of the form,

$$\lambda^*(t) = \mu + \int_0^t \phi(t-u) \mathrm{d}N(u),$$

for some $\mu > 0$ and $\phi : (0, \infty) \to [0, \infty)$ which are called the background intensity and excitation function respectively. Assume that $\phi(\cdot) \neq 0$. Such a process $N(\cdot)$ is a Hawkes process.[a] Examples of excitation functions $\phi(\cdot)$,

- $\phi(t) = \alpha\beta \exp(-\beta t) 1_{\{t \geq 0\}}$
- $\phi(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\nu)^2}{2\sigma^2}\right)$

[a] Patrick J. Laub, Thomas Taimre, and Philip K. Pollett. *Hawkes Processes*. 2015.

Conditional Intensity

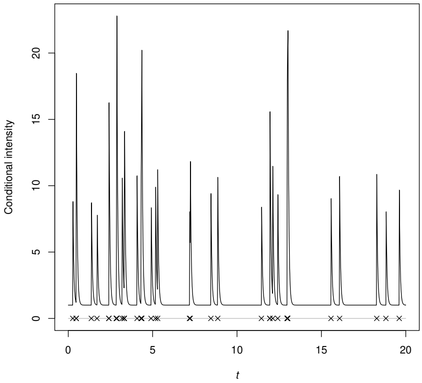


Figure: $\lambda^*(t) = 1 + 0.5 \cdot 20 \exp(-20t) 1_{\{t \geq 0\}}$.

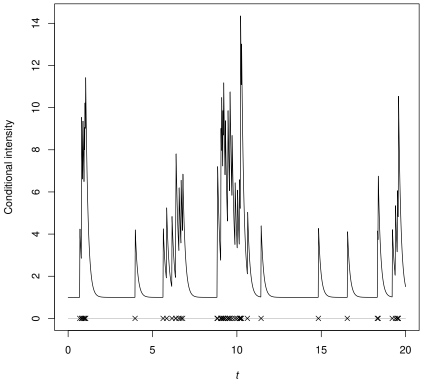
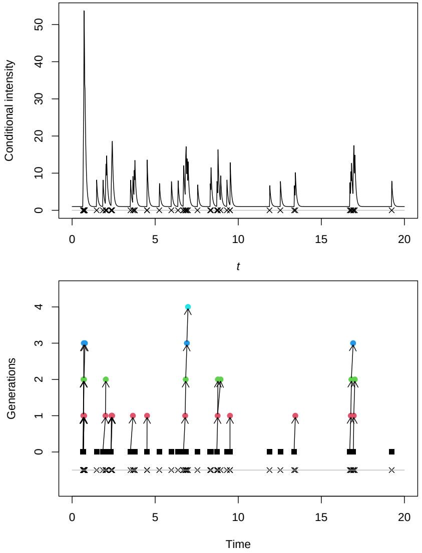


Figure: $\lambda^*(t) = 1 + 0.5 \cdot 7 \exp(-7t) 1_{\{t \geq 0\}}$.

Immigration–Birth Representation



Immigration–Birth Representation

Recall the generic conditional intensity,

$$\lambda^*(t) = \mu + \int_0^t \phi(t-u)\mathrm{d}N(u).$$

The Hawkess process can be seen as initial inflow of *immigrants* followed by generation of *children*,

- *the immigrants* arrive with an homogeneous poisson, Pois(λ).
- *the children*: An immigration at T_i generates children according to an inhomogeneous Poisson process of intensity $\lambda(t) = \phi(t - T_i)$.

The children can in turn also independently generate children. The superposition of all the branching processes is a Hawkes process of conditional intensity, $\lambda^*(t)$. E.g.,

$$\lambda^*(t) = 1 + 0.5 \cdot 15\exp(-15t)1_{\{t\geq 0\}}$$

Average Intensity

Under the assumption of a ergodic and stationary process, let $\bar{\lambda} = \mathbb{E}\left[\lambda(t)\right]$, then,

$$\hat{\lambda} = \mathbb{E}\left[\lambda(t)\right] = \mathbb{E}\left[\mu + \int_0^t \phi(t-s)\mathrm{d}N(s)\right] = \mu + \int_0^t \phi(t-s)\mathbb{E}[\mathrm{d}N(s)]$$

Average Intensity

Under the assumption of a ergodic and stationary process, let $\bar{\lambda} = \mathbb{E}\left[\lambda(t)\right]$, then,

$$\hat{\lambda} = \mathbb{E}\left[\lambda(t)\right] = \mathbb{E}\left[\mu + \int_0^t \phi(t-s)\mathrm{d}N(s)\right] = \mu + \int_0^t \phi(t-s)\mathbb{E}[\mathrm{d}N(s)]$$

How to calculate $\mathbb{E}[\mathrm{d}N(s)]$?

$$\lambda(s) = \lim_{h \downarrow 0} \frac{\mathbb{E}[N(s+h) - N(s) \mid \mathcal{F}(s)]}{h} = \frac{\mathbb{E}[\mathrm{d}N(s) \mid \mathcal{F}(s)]}{\mathrm{d}s}$$

Average Intensity

Under the assumption of a ergodic and stationary process, let $\bar{\lambda} = \mathbb{E}\left[\lambda(t)\right]$, then,

$$\hat{\lambda} = \mathbb{E}\left[\lambda(t)\right] = \mathbb{E}\left[\mu + \int_0^t \phi(t-s)\mathrm{d}N(s)\right] = \mu + \int_0^t \phi(t-s)\mathbb{E}[\mathrm{d}N(s)]$$

How to calculate $\mathbb{E}[\mathrm{d}N(s)]$?

$$\lambda(s) = \lim_{h \downarrow 0} \frac{\mathbb{E}[N(s+h) - N(s) \mid \mathcal{F}(s)]}{h} = \frac{\mathbb{E}[\mathrm{d}N(s) \mid \mathcal{F}(s)]}{\mathrm{d}s}$$

Take expectation and use tower property,

$$\bar{\lambda} = \mathbb{E}\left[\lambda(s)\right] = \frac{\mathbb{E}[\mathbb{E}[\mathrm{d}N(s) \mid \mathcal{F}(s)]]}{\mathrm{d}s} = \frac{\mathbb{E}[\mathrm{d}N(s)]}{\mathrm{d}s}$$

Average Intensity

Under the assumption of a ergodic and stationary process, let $\bar{\lambda} = \mathbb{E}\left[\lambda(t)\right]$, then,

$$\hat{\lambda} = \mathbb{E}\left[\lambda(t)\right] = \mathbb{E}\left[\mu + \int_0^t \phi(t-s)\mathrm{d}N(s)\right] = \mu + \int_0^t \phi(t-s)\mathbb{E}[\mathrm{d}N(s)]$$

How to calculate $\mathbb{E}[\mathrm{d}N(s)]$?

$$\lambda(s) = \lim_{h \downarrow 0} \frac{\mathbb{E}[N(s+h) - N(s) \mid \mathcal{F}(s)]}{h} = \frac{\mathbb{E}[\mathrm{d}N(s) \mid \mathcal{F}(s)]}{\mathrm{d}s}$$

Take expectation and use tower property,

$$\bar{\lambda} = \mathbb{E}\left[\lambda(s)\right] = \frac{\mathbb{E}[\mathbb{E}[\mathrm{d}N(s) \mid \mathcal{F}(s)]]}{\mathrm{d}s} = \frac{\mathbb{E}[\mathrm{d}N(s)]}{\mathrm{d}s}$$

Hence we can write,

$$\bar{\lambda} = \mu + \int_0^t \bar{\lambda}\phi(t-s)\mathrm{d}s = \mu + \bar{\lambda}\int_0^\tau \phi(\tau)\mathrm{d}\tau,$$

$$\tau = t - s$$

Average Intensity

Under the assumption of a ergodic and stationary process, let $\bar{\lambda} = \mathbb{E}\left[\lambda(t)\right]$, then,

$$\hat{\lambda} = \mathbb{E}\left[\lambda(t)\right] = \mathbb{E}\left[\mu + \int_0^t \phi(t-s)\mathrm{d}N(s)\right] = \mu + \int_0^t \phi(t-s)\mathbb{E}[\mathrm{d}N(s)]$$

How to calculate $\mathbb{E}[\mathrm{d}N(s)]$?

$$\lambda(s) = \lim_{h \downarrow 0} \frac{\mathbb{E}[N(s+h) - N(s) \mid \mathcal{F}(s)]}{h} = \frac{\mathbb{E}[\mathrm{d}N(s) \mid \mathcal{F}(s)]}{\mathrm{d}s}$$

Take expectation and use tower property,

$$\bar{\lambda} = \mathbb{E}\left[\lambda(s)\right] = \frac{\mathbb{E}[\mathbb{E}[\mathrm{d}N(s) \mid \mathcal{F}(s)]]}{\mathrm{d}s} = \frac{\mathbb{E}[\mathrm{d}N(s)]}{\mathrm{d}s}$$

Hence we can write,

$$\bar{\lambda} = \mu + \int_0^t \bar{\lambda}\phi(t-s)\mathrm{d}s = \mu + \bar{\lambda}\int_0^\tau \phi(\tau)\mathrm{d}\tau,$$

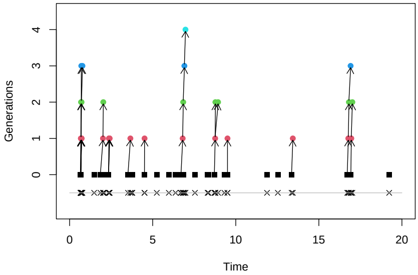
$$\tau = t - s$$

Stationarity

We must require $R_0 < 1$,

$$\bar{\lambda} = \frac{\mu}{1 - R_0}, \qquad R_0 = \int_0^\infty \phi(\tau)\mathrm{d}\tau$$

Correlation - A Cluster Ratio Approach



Clustering Ratio for a Hawkes Process

It can be shown that the cluster ratio for a Hawkes process is,

$$\rho = \frac{1}{(1 - R_0)^2} \geq 1$$

Notice, $\rho \perp\!\!\!\perp \mu$ and depends only on ϕ through,

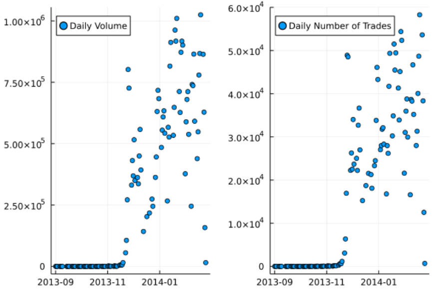
$$\int_0^\infty \phi(\tau) \mathrm{d}\tau$$

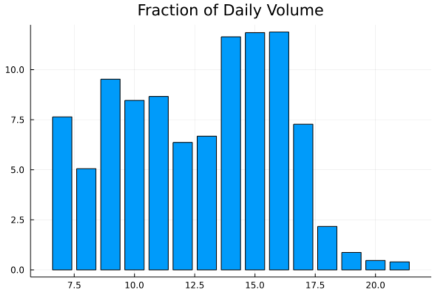
Further notice the implications of the limits,

$$R_0 \to 0 \qquad \rho \to 1$$
$$R_0 \to \infty \qquad \rho \to \infty$$

When calibrated to financial markets $0 << \rho < 1$.

High Frequency Data - An Example[3]





[3] Markwick, *Modelling Microstructure Noise Using Hawkes Processes*.

High Frequency Data - Quadratic Variation

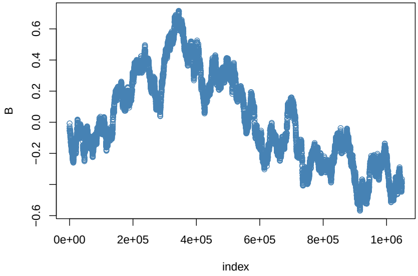
Bounded Quadratic Variation

Partition an interval $[0,T]$ into $\#\Delta$ sub-intervals, and let $|\Delta|$ denote the largest. Then the quadratic variation of Brownian motion is,

$$[B]_t = \lim_{|\Delta|\to 0} \sum_{i=1}^{\#\Delta} \left|B_{t_i} - B_{t_{i-1}}\right|^2,$$

That is, $[B]_T = T$.

High Frequency Data - Quadratic Variation



Bounded Quadratic Variation

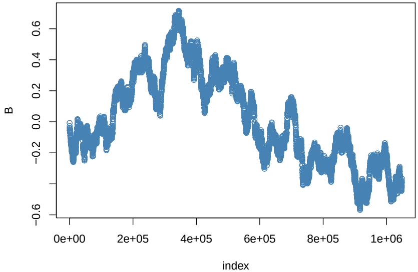
Partition an interval $[0, T]$ into $\#\Delta$ sub-intervals, and let $|\Delta|$ denote the largest. Then the quadratic variation of Brownian motion is,

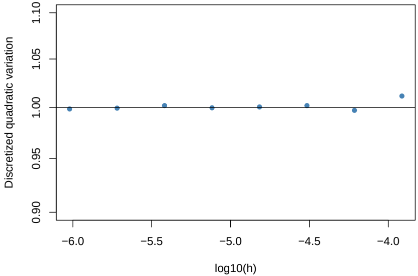
$$[B]_t = \lim_{|\Delta|\to 0} \sum_{i=1}^{\#\Delta} \left| B_{t_i} - B_{t_{i-1}} \right|^2,$$

That is, $[B]_T = T$.

Consider the specific example with $T = 1$ and an experiment where we sub-sample. What is going to happen as $|\Delta| \to 0$. In the end we will have, $\#\Delta = 2^{20}$.

High Frequency Data - Quadratic Variation





Bounded Quadratic Variation

Partition an interval $[0,T]$ into $\#\Delta$ sub-intervals, and let $|\Delta|$ denote the largest. Then the quadratic variation of Brownian motion is,

$$[B]_t = \lim_{|\Delta|\to 0} \sum_{i=1}^{\#\Delta} \left| B_{t_i} - B_{t_{i-1}} \right|^2 ,$$

That is, $[B]_T = T$.

Consider the specific example with $T = 1$ and an experiment where we sub-sample. What is going to happen as $|\Delta| \to 0$. In the end we will have, $\#\Delta = 2^{20}$.

High Frequency Data - An Example

Signature Plot

Consider a period $[0, T]$ at a specific scale $\tau > 0$, then the realized volatility is,

$$\widehat{C}(\tau) = \frac{1}{T} \sum_{n=0}^{T/\tau} |X((n+1)\tau) - X(n\tau)|^2 \quad (0.1)$$

Compare this with the definition of quadratic variation, where $\#\Delta$ is the number of sub-intervals,

$$[B]_t = \lim_{|\Delta| \to 0} \sum_{i=1}^{\#\Delta} \left| B_{t_i} - B_{t_{i-1}} \right|^2,$$

i.e. the realized volatility is the observed quadratic variation.

High Frequency Data - An Example

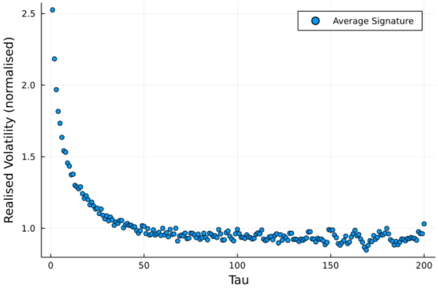


Figure: Microstructure Noise in the Futures Data.

Signature Plot

Consider a period $[0, T]$ at a specific scale $\tau > 0$, then the realized volatility is,

$$\widehat{C}(\tau) = \frac{1}{T} \sum_{n=0}^{T/\tau} |X((n+1)\tau) - X(n\tau)|^2 \quad (0.1)$$

Compare this with the definition of quadratic variation, where $\#\Delta$ is the number of sub-intervals,

$$[B]_t = \lim_{|\Delta| \to 0} \sum_{i=1}^{\#\Delta} \left| B_{t_i} - B_{t_{i-1}} \right|^2,$$

i.e. the realized volatility is the observed quadratic variation.

High Frequency Data - A Hawkess Price Model

Introducing the Model

Let $\{X(t) : t \in [0, T]\}$ denote the price t with $X(0) = 0$ for simplicity, then we introduce,

$$X(t) = N_1(t) - N_2(t),$$

where $N_i(t) \overset{iid}{\sim}$ Pois(λ) and $N_1(t)$, $N_2(t)$ represent the sum of positive and negative jumps respectively. The couple $\{N_1(t), N_2(t)\}$ is a bivariate Hawkess process and for simplicity we assume that the processes are *purely mean reverting* i.e.,

$$\lambda_1(t) = \mu + \int_{-\infty}^{t} \varphi(t-s)\mathrm{d}N_2(s)$$

$$\lambda_2(t) = \mu + \int_{-\infty}^{t} \varphi(t-s)\mathrm{d}N_1(s)$$

Choice of kernel

For the specific model, we consider the right-sided exponential function,

$$\varphi(t) = R_0\beta e^{-\beta t}1_{\mathbb{R}+}(t)$$

and as shown previously, we require

$$1 > R_0 = \left[-R_0 e^{-\beta t}\right]_{t=0}^{t=\infty} = R_0$$

Diffusion on Large Scale

It can be shown that the model diffuses on a large scale time scale, i.e. if we let $T \to \infty$, then

$$\lim_{T\to+\infty} \frac{1}{\sqrt{T}} X(tT) \overset{(d)}{=} \sqrt{2\lambda}B(t), \quad t \in [0, 1],$$

where $\sqrt{2\lambda}$ is the macroscopic volatility.

Signature With Model and Data

Signature Plot under the Model

Consider $\{X(t) : [0, T]\}$ and *wlog* let $X(0) = 0$, then the mean signature plot is,

$$C(\tau) = \mathbb{E}[\widehat{C}(\tau)] = \frac{1}{\tau}\mathbb{E}\left[X(\tau)^2\right]$$

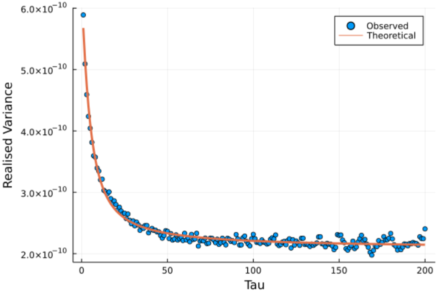
If $1 < R_0$, then we have a closed form expression for $C(\tau)$,

$$C(\tau) = \Lambda\left(\kappa^2 + \left(1 - \kappa^2\right)\frac{1 - e^{-\gamma\tau}}{\gamma\tau}\right) \qquad (0.2)$$

where

$$\Lambda = \frac{2\mu}{1 - \|\varphi\|_1}, \quad \kappa = \frac{1}{1 + R_0},$$
$$\gamma = \beta\left(R_0 + 1\right)$$

Signature With Model and Data



Signature Plot under the Model

Consider $\{X(t) : [0,T]\}$ and *wlog* let $X(0) = 0$, then the mean signature plot is,

$$C(\tau) = \mathbb{E}[\widehat{C}(\tau)] = \frac{1}{\tau}\mathbb{E}\left[X(\tau)^2\right]$$

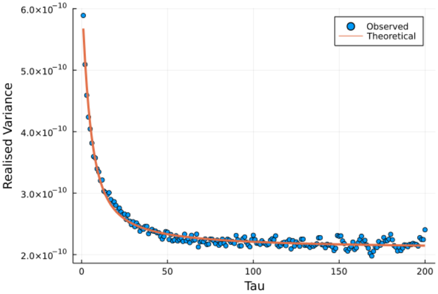
If $1 < R_0$, then we have a closed form expression for $C(\tau)$,

$$C(\tau) = \Lambda\left(\kappa^2 + \left(1-\kappa^2\right)\frac{1-e^{-\gamma\tau}}{\gamma\tau}\right) \qquad (0.2)$$

where

$$\Lambda = \frac{2\mu}{1-\|\varphi\|_1}, \quad \kappa = \frac{1}{1+R_0},$$
$$\gamma = \beta\left(R_0+1\right)$$

Signature Model Fitting and Interpretation



Obtained Model Parameters

- $\mu = 0.24402$
- $R_0 = 0.74179$
- $\beta = 0.19569$

Fitting the model

Recall that the realized signature plot over the period is,

$$\widehat{C}(\tau) = \frac{1}{T} \sum_{n=0}^{T/\tau} |X((n+1)\tau) - X(n\tau)|^2$$

thus we could use a regression estimator,
$$\widehat{\theta}_{\text{reg}} = \text{Argmin}_\theta \, |\widehat{C}(\tau) - C(\tau)|^2$$

Interpretation of Parameters

- $R_0 = 0.75$ thus large excitement with each jump.
- $\beta = 0.2 \quad \implies$ it on avg. takes $\frac{1}{0.2} \approx 5$ sec. from a downstick to an upstick.

Laub, Patrick J., Thomas Taimre, and Philip K. Pollett. *Hawkes Processes*. 2015.

— .*Hawkes Processes*. 2015.

Markwick, Dean. *Modelling Microstructure Noise Using Hawkes Processes*. URL: https://dm13450.github.io/2022/05/11/modelling-microstructure-noise-using-hawkes-processes.html.