High-Frequency Market Microstructure Noise Hawke's Process Perspective

Rishi Dey Chowdhury (MB2320) | Quantitative Finance | Assignment 2

M.Stat. 2nd Year, Indian Statistical Institute, Kolkata

Introduction

The exponential rise of electronic markets, automated trading and compute power has lead to industries shifting gear from end of the day trading at the closing price to high-frequency low-latency trading within fraction of seconds. The industry standard for tick-to-trigger time stands somewhere around $30 - 60$ nanoseconds. With the availability of a huge number of quality high frequency data, there is a fast growing literature devoted to the modelling of intradaily asset prices behaviour. Since Bachelier and the seminal work of Black and Scholes, most popular models at a coarse time scale – for daily data, say – are Brownian diffusions. In particular, diffusion models aim at describing more or less faithfully the volatility dynamics, characterized by stylized facts such as volatility clustering or leverage effect. A key issue that naturally emerges when one studies high frequency data is the problem of improving volatility estimation over a given time period, thanks to the massive amount of data available at such scales nowadays. The discrete nature of time trade arrivals and of price variations (prices are point processes living on a tick grid), the presence of so-called microstructure noise (described as strong mean reversion effects at small scales) makes this question highly non trivial. At a very high frequency, prices variations are also characterized by well documented stylized facts like the signature plot behaviour.

High Frequency Volatility - Microstructure Noise

If $X(t)$ stands for the price of some asset at time t (defined indifferently as the last traded price or the mid-price between best bid and best offer in the order book), the signature plot can be defined from the quadratic variation of $X(t)$ over a time period $[0, T]$ at a scale $\tau > 0$ – the so-called realized volatility - as

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

The microstructure noise effect manifests through an increase of the observed daily variance when one goes from large to small scales i.e. in the limit $\tau \to 0$.

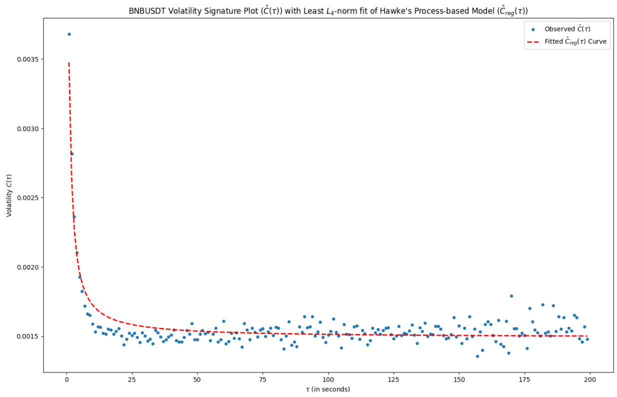

Figure 1: *Signature volatility plot of the observed BNBUSDT closing price per second trading on Binance Cryptocurrency Exchange on 30^{th} October 2024 along with the fitted mutually exciting Hawkes Process' theoretical signature plot.*

This behaviour is different from what one would expect if the data were sampled from a Brownian diffusion, in which case the function above should be flat. From the perspective of statistical estimation, this leads to a simple paradox. On one side, the smaller τ , the larger is the dataset that can be use to estimate the volatility. However, how should one be using high-frequency data in order to obtain better estimates of the volatility, since the realized volatility is not stable as τ decreases.

In the literature the most popular approaches attempts to model microstructure noise with the concept of latent price. One starts with a Brownian diffusion $X(t)$ defined as an efficient price, which is latent, in the sense that it cannot be observed directly. Instead the practitioner has only access to noisy version $\widetilde{X}(t)$ of $X(t)$ that accounts for microstructure noise. Below, we have the additive version of the microstructure noise model,

$$\widetilde{X}(n\tau) = X(n\tau) + \xi_{n,\tau}$$

where the microstructure noise term $\xi_{n,\tau}$ satisfies $\mathbb{E}(\xi_{n,\tau}) = 0$ for obvious identifiability conditions. The goal is then to separate the noise from the true signal $X(t)$, from which a classical volatility estimator can be performed. Whereas the above representation produces an elegant pilot model to describe microstructure noise effects at the scale of a few minutes, it cannot faithfully reproduce the data as they are observed on a microscopic scale of a few seconds: for instance, the discreteness of price changes is left out and the mathematical artifact of forcing $\hat{C}(\tau)$ to explode when $\tau \to 0$. Here we will see a tick-by-tick model by means of marked point process with appropriate stochastic intensities, we are able to control its features at all scale. We will rely on multivariate Hawkes Processes associated with positive and negative jumps of the asset prices.

Methodology

We start with two point processes $N_1(t)$ and $N_2(t)$ for $t \in [0, T]$ that represent respectively the sum of positive and negative jumps of some asset price $X(t)$ over some time horizon $[0, T]$,

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

If $N_1(t)$ and $N_2(t)$ are two independent Poisson processes with intensity μ, it is easy to show that the model diffuses at large scale, i.e., when $T \to \infty$, by introducing the scaling factor $T^{1/2}$, we obtain the following limit in distribution:

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

where $\sqrt{2\mu}$ is the diffusive or macroscopic volatility, that accounts for the activity of negative and positive jumps, hence the factor 2 in the limit. According to equation of $\widehat{C}(\tau)$, the corresponding mean signature plot is flat: for all $\tau > 0$, we have $\mathbb{E}[\widehat{C}(\tau)] = 2\mu$. In order to account for the previously reported noise microstructure features, intuitively and as confirmed by empirical observations, one has to introduce some mean reversion in the small scales, while ensuring that this mean reversion effect vanishes on large scale. This can be naturally done within the context of (multivariate) Hawkes process as follows.

We extend in a first step the time horizon $[0, T]$ over the whole real line $\mathbb{R} = (-\infty, +\infty)$. Let $\lambda_i(t)$ for $t \in \mathbb{R}$ be the stochastic intensities of two counting processes $N_i(t), i = 1, 2$, such that at time t :

$$\lambda_i(t) = \lim_{\Delta \to 0} \Delta^{-1} \mathbb{E}\left[N_i(t + \Delta) - N_i(t) \mid \mathcal{F}_t\right]$$

where $\mathcal{F}_t$ stands for the filtration generated by the history of the processes $N_1(t), N_2(t)$. The bivariate process $\{N_1(t), N_2(t)\}$ is a linear Hawkes process if $N_1(t)$ and $N_2(t)$ have no common jumps and if there exist four nonnegative functions $\{\varphi_{ij}\}_{i,j=1,2}$ such that

$$\lambda_i(t) = \mu_i + \int_{-\infty}^{t} \varphi_{ii}(t - u) dN_i(u)$$

$$+ \int_{-\infty}^{t} \varphi_{ij}(t - u) dN_j(u)$$

The so-obtained process can be shown to be well defined and to admit a version with stationary increments under the stability condition

$$\text{all the eigenvalues of the matrix } \left\{ \left\| \varphi_{ij} \right\|_1 \right\} \text{ are } < 1,$$

where $\|\varphi\|_1 = \int_{\mathbb{R}} \varphi(t) dt$. Mean reversion can be translated by the fact that the more $X(t)$ goes up, the greater the intensity $\lambda_2(t)$ and conversely, the more $X(t)$ goes down, the greater the intensity $\lambda_1(t)$. This leads to the following simplified version of previous model (where only mean-reverting terms were kept):

$$\lambda_1(t) = \mu + \int_{-\infty}^{t} \varphi(t - s) dN_2(s)$$

$$\lambda_2(t) = \mu + \int_{-\infty}^{t} \varphi(t - s) dN_1(s)$$

where μ is an exogenous intensity and $\varphi(t)$ a positive kernel which is causal (i.e., $\text{Supp}(\varphi) \subset \mathbb{R}^+$). Equations for $\lambda_1(t)$ and $\lambda_2(t)$ define two mutually exciting point processes that are stationary and stable under the condition $\|\varphi\|_1 < 1$.

A simple and natural choice for φ is a right-sided exponential function:

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

where $\alpha, \beta > 0$ are such that

$$\|\varphi\|_1 = \frac{\alpha}{\beta} < 1$$

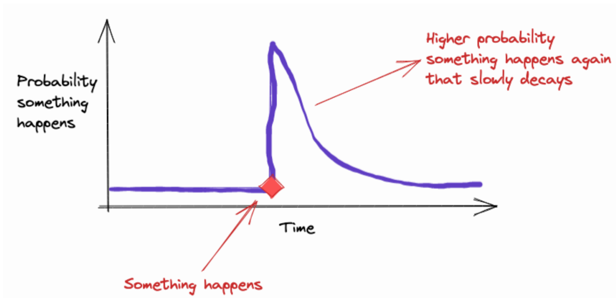

Figure 2: *An upward jump increases the intensity/chances of a downward jump happening for a short-lived time after the previous jump. This intensity gradually fades away due to exponential decay. This is why it is also called* ***Mutually Exciting Hawke's Process****.*

Signature plot

We are interested only in $X(t)$ on $[0, T] \subset \mathbb{R}_+$and for simplicity and without loss of generality, we set $X(0) = 0$. From the expression of $\widehat{C}(\tau)$, the mean signature plot can be written as

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

where we can write,

$$\frac{1}{\tau}\mathbb{E}\left[|X((n+1)\tau) - X(n\tau)|^2\right] = \frac{1}{\tau}\mathbb{E}\left[X(\tau)^2\right]$$

If $\varphi(t)$ is defined as the exponential kernel we saw before, then it is possible to obtain a closed form for the mean signature plot,

Under the stability condition $\|\varphi\|_1 = \frac{\alpha}{\beta} < 1$, we have

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

where

$$\Lambda = \frac{2\mu}{1-\|\varphi\|_1}, \quad \kappa = \frac{1}{1+\|\varphi\|_1}, \quad \text{and } \gamma = \alpha + \beta$$

We can see in particular a cross-over from the microstructural variance

$$V_0 = \mathbb{E}[\widehat{C}(0)] = \Lambda = 2\mathbb{E}\left(\lambda_i\right),$$

to the diffusive variance

$$V_\infty = \mathbb{E}[\widehat{C}(\infty)] = \Lambda\kappa^2$$

Simulation and Parameter estimation

We consider the exponential kernel in the Hawke's process we saw before. There are 3 parameters, namely : $\theta = (\mu, \alpha, \beta)$.

Simulation of this process on an interval $[0, T]$ can be performed using the thinning algorithm. It basically consists in simulating on $[0, T]$ a standard Poisson process with an intensity M large enough such that it satisfies the following condition:

$$\lambda_1(t) < M \text{ and } \lambda_2(t) < M, \text{ for all } t \in [0, T]$$

A thinning procedure is then applied to each jump of the obtained process from the first one to the last one allowing to either reject the point (with probability $\frac{M-\lambda_1(t)-\lambda_2(t)}{M}$) or mark it as a jump of $N_1(t)$ with probability $\frac{\lambda_1(t)}{M}$ or of $N_2(t)$ with probability $\frac{\lambda_2(t)}{M}$, where t is the time of the considered jump.

The estimation of the parameters can be done using the ability of the model to reproduce the mean signature plot, the parameters can be estimated using a best fit of the realized signature plot. The realized signature plot over $[0, T]$ is defined as

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

[0] [1] Indeed, since only the increments of the processes $N_i(t)$ come into play, we may (and will) assume that $N_1(t) = N_2(t) = 0$ hence $X(0) = 0$.

and regression estimator $\widehat{\theta}_{\text{reg}}$ is then naturally given by

$$\widehat{\theta}_{\text{reg}} = \text{Argmin}_\theta\,|\widehat{C}(\tau) - C(\tau)|^2$$

where $C(\tau)$ is the theoretical expected signature plot defined before.

For regression estimates $\widehat{\theta}_{\text{reg}}$ the minimization must be performed under the constraints

$$\mu > 0, \quad \alpha > 0, \quad \beta > 0$$

$$\text{and the stability condition : } \frac{\alpha}{\beta} < 1$$

A fitted signature plot to the BNBUSDT price data can be seen in figure 1.

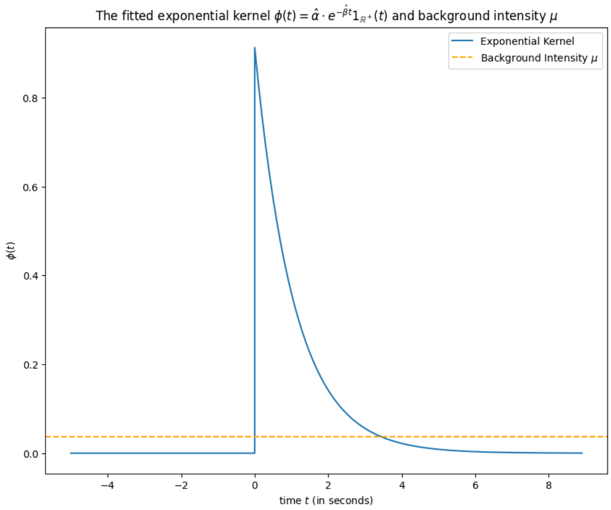

Figure 3: *The estimated regression parameters on the BNBUSDT data $\widehat{\theta}_{\text{reg}}$ has $\hat{\alpha}$ and $\hat{\beta}$ which gives us the fitted kernel as shown above. We can see how a jump increases the intensity of the other jump drastically which gradually fades away and drops to almost half within 1 second.*

Real-Life Data Application

Data

I collected ticker orderbook data with a frequency of 1 tick per second from **Binance** cryptocurrency exchange for BNBUSDT on 30^{th} October 2024 shown in figure 4. It contains a total of 86400 closing prices for each second. Cryptocurrency exchanges are functional 24×7 hence we get the price data for the entire 24 hours of 30^{th} October. The tick size for the BNBUSDT crypto pair is 0.1$.

Exploratory Data Analysis

One of the important questions before we try to fit a point process is to know how the price jumps between $X(t) \to X(t+1)$ are distributed. We have

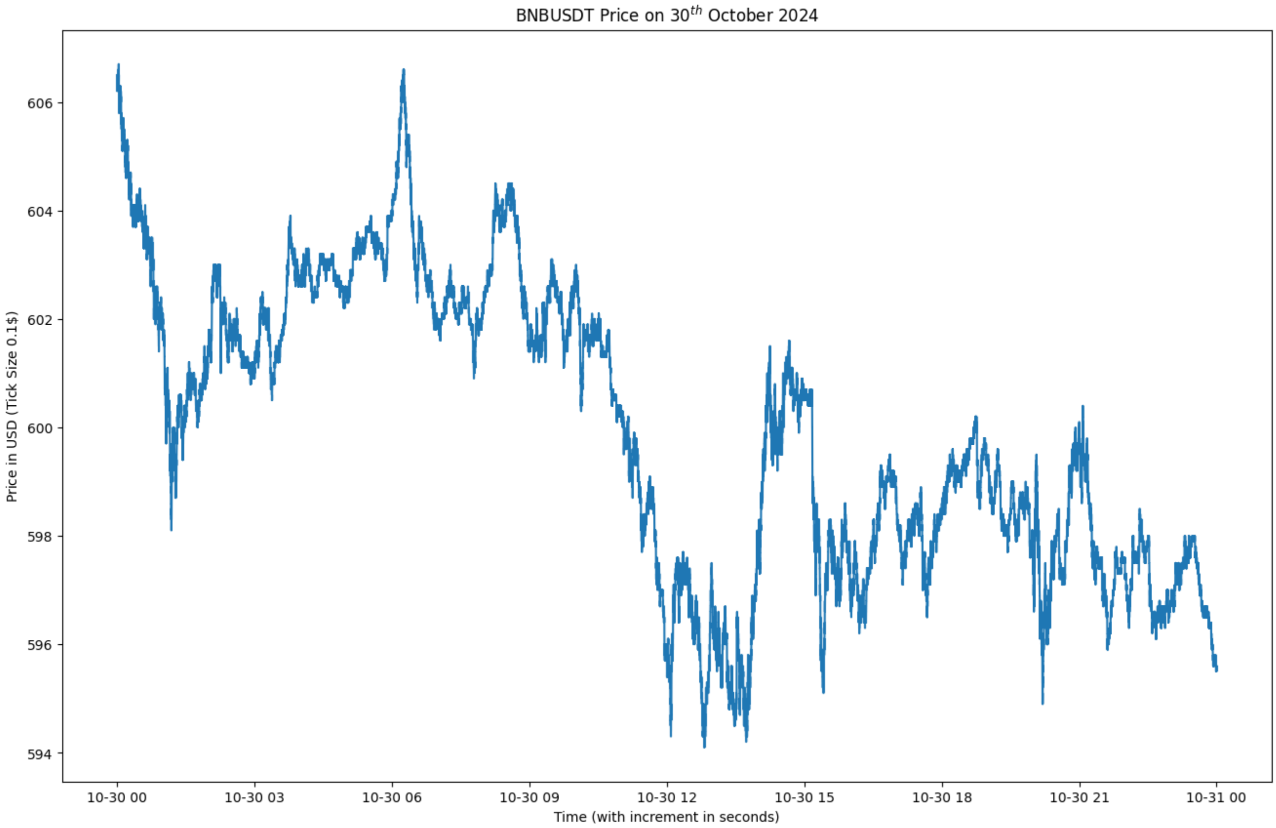

Figure 4: *The per second closing price of the BNBUSDT ticker on* 30th *October 2024 from 0000 hours 00 seconds to 2359 hours 59 seconds*

shown result from our dataset in figure 5. We observe the Poisson-like distribution for both N_1 and N_2 jumps. We might observe ≥ 2 jumps in one second but the probability is extremely low similar to Poisson process. It seems that the apriori probability of observing jump of anyone kind is roughly equal.

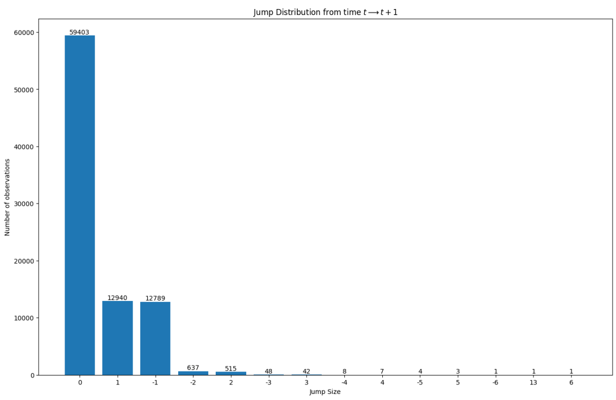

Figure 5: *Jump Size distribution from* $X(t) \longrightarrow X(t+1)$ *shows almost symmetric number of* N_1 *and* N_2 *jumps.* 1 *indicates an* N_1 *jump whereas* -1 *indicates an* N_2 *jump.*

Apart from looking at the count of the number of jumps which is roughly a Poisson Distribution as seen in figure 5, we also take a look at the inter-arrival times of N_1 and N_2 jumps respectively in 6. It strongly resembles a exponential distribution giving us a good idea of the jumps being distributed as a Poisson Process. But as seen in the methodology section if we used two independent constant intensity poisson processes to model $N_1(t)$ and $N_2(t)$, we won't be able to replicate the micro-structural level mean-reversion and the signature plot which would turn out to be flat. Hence, the choice to fit the Hawke's Process based model seems justified.

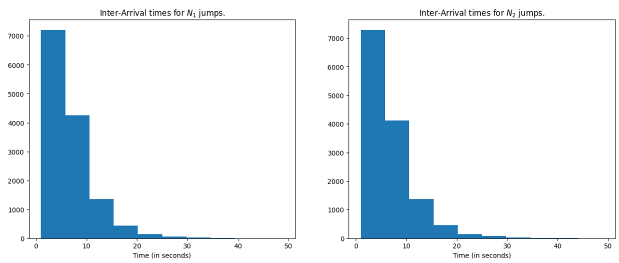

Figure 6: *Distribution of Inter-arrival times of upward* (N_1) *and downward* (N_2) *jumps respectively resembles an Exponential distribution.*

Hawke's Process Price Model

We here discuss about the fitted mutually exciting Hawke's process based price model. Although *MLE*

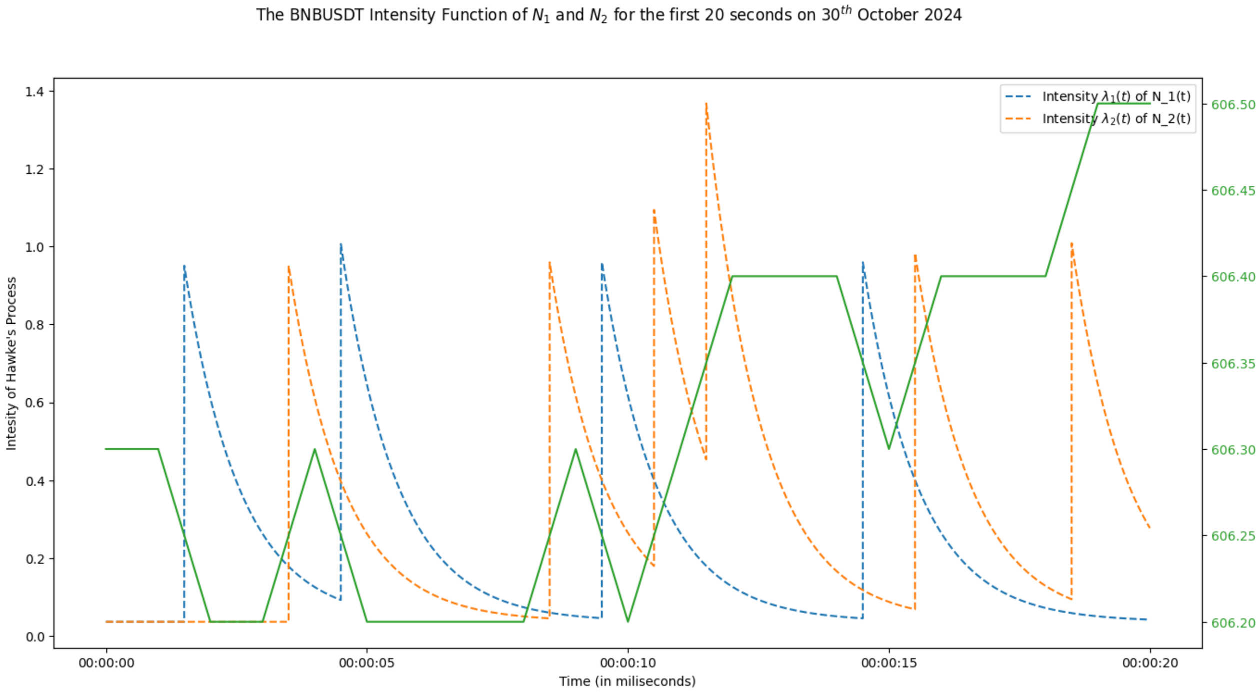

Figure 7: *The underlying intensity function of $N_1(t)$ and $N_2(t)$ from the fitted Mutually Exciting Hawke's Process for BNBUSDT prices in the first* 20 *seconds of* 30^{th} *October 2024. A* N_1 *jump increase* λ_2 *which gradually fades away and vice-versa.*

could be used to estimate such a model the unavailability of the exact arrival times of the jumps (i.e. there can be multiple jumps in one second) and the granularity of the data being only upto one second implies the arrival times t happening between s seconds and $s+1$ seconds would be reported as the nearest ceiling second i.e. $s+1$. So, determining the exact number of jumps of both kind becomes difficult as well as their arrival times. Moreover, the computation time for MLE-based estimate is quite high and somewhat unstable since we don't have direct access to the underlying point process. The regression based estimator is fast and is much more robust when fitting to observed data.

In table 1, we can see the estimated parameters of the Hawke's process price model which we have fit using the **Nelder-Mead** algorithm to minimize the squared distance between the estimated theoretical signature plot and the observed signature plot with $\tau = 1, \ldots, 200$. The $\frac{1}{\hat{\beta}} \approx 1$ seconds is roughly the half-life time of the effect of one jump increasing the intensity of the other jump happening. $\frac{\hat{\alpha}}{\hat{\beta}} \approx 0.98 < 1$ shows that the stability condition is met and the sudden increase in intensity of the other jump when a jump happens is very high. $\hat{\mu}$ is the background intensity of a N_1 (or N_2) jump happening which is quite low.

The fitted signature plot can be seen in figure 1 and the fitted exponential kernel can be observed in figure 3. Figure 7 shows how both the intensity

$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$
0.03733	0.9133	0.9316

Table 1: *The estimated parameters of the mutually exciting Hawke's process.*

functions behave with price changes. Notice how a N_1 jump drastically increases the λ_2 intensity of N_2 and vice-versa. This captures the mean-reversion behaviour and the signature plot to effectively model the market microstructure at such high-frequency scenario.

Model Performance

Signature Plot The fitted Hawke's process price model replicates the observed signature plot with a mean-squared error of **0.0505**. We compare the model we have fit to the data on 31^{st} October 2024 as shown in figure 8. We observe a mean squared error of **0.0858** compared to the best fit's mean squared error being **0.0612**. The best fit's estimated parameter on 31^{st} October 2024 is reported in table 2.

$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$
0.0534	0.9733	1.749

Table 2: *The estimated parameters of the mutually exciting Hawke's process with data on* 31^{st} *October 2024.*

The estimated parameters of our model are roughly close to those of the new best model on 31^{st} October

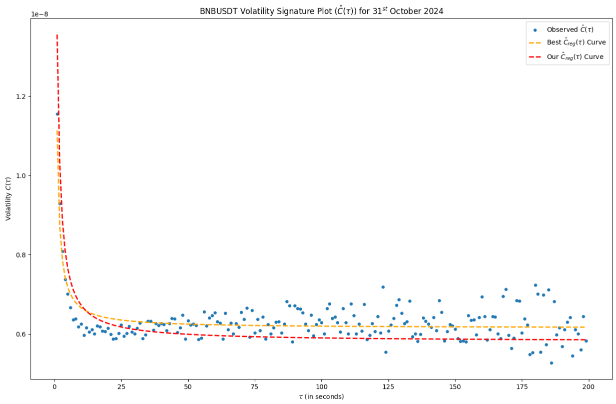

Figure 8: *The signature plot of the observed data on 31^{st} October 2024 with our previous fitted signature plot compared with the best fit.*

2024. The signature plot also shows how our model performs decently in replicating the microstructure noise. In production, usually we adapt and re-calibrate our model online with more and more data at high frequency and the differences we see here will be far more less.

Jump Prediction Since, we have a calibrated model with varying underlying intensities we can see how well the model is able to predict the jumps in our data which will give us confidence about the model being good at capturing the microstructure level properties of the data. $N_1(t)$ and $N_2(t)$ are two independent inhomogeneous Poisson processes with varying intensity functions $\lambda_1(t)$ and $\lambda_2(t)$. We can use this to model the probability of a jump happening given the prices we have observed till now. If we are at time t seconds, we can find the probabilities of being at the same price at time (p_0), higher price (p_1) or lower price (p_2) at time $t + 1$ as follows,

$$N_1(t+1) - N_1(t) \sim Pois(\Lambda_1 = \int_t^{t+1} \lambda_1(s)ds)$$

$$N_2(t+1) - N_2(t) \sim Pois(\Lambda_2 = \int_t^{t+1} \lambda_2(s)ds)$$

$$\begin{aligned} p_0 &= P[N_1(t+1) - N_1(t) = N_2(t+1) - N_2(t)] \\ &= e^{-(\Lambda_1+\Lambda_2)} \sum_{k=0}^{\infty} \frac{(\Lambda_1 \Lambda_2)^{k!}}{k!^2} \\ p_1 &= P[N_1(t+1) - N_1(t) > N_2(t+1) - N_2(t)] \\ &= (1 - p_0) \cdot \frac{\Lambda_1}{\Lambda_1 + \Lambda_2} \\ p_2 &= P[N_1(t+1) - N_1(t) < N_2(t+1) - N_2(t)] \\ &= (1 - p_0) \cdot \frac{\Lambda_2}{\Lambda_1 + \Lambda_2} \end{aligned}$$

So for each time t, the predicted probabilities were found for time $t + 1$ on 30^{th} October. Since, the probability of p_0 is usually very high due to most of the jumps being of size 0 as seen in figure 5, we take estimate that a jump occurs when $p_0 < 0.8$. In all these data points our model accurately predicts the outcome of the next ticker price's jump direction with **64.14%** accuracy only using the Hawke's process-based price model which is good given the financial context.

Hence, we conclude the model fits the data well and appropriately captures the volatility from fine to coarse scale. It also replicates the jumps in the dataset decently good.

Conclusion

Studying these high-frequency compatible stochastic models for limit orderbook, trades and price modelling are of high importance specially when the markets are shifting towards low-latency trading solutions. They reveal several insights about the price dynamics and holds a chance to lead to a profitable trading strategy. The estimation of volatility at such high frequency is important for efficiently pricing options. There exists several other variations of Hawke's Process taking into account the entire orderbook and trades. Recent trends towards Machine Learning and Deep Learning has also lead to non-linear modelling of the intensity function. We can combine external informations like US Election results, news about frauds, etc to simulate jumps using Hawke's Process to enhance the detection of large jumps or crashes when certain jump intensities are high. Some interesting follow up exploration can extend the method to take into account volume imbalance and trade information as well as news embeddings. Certain metrics like probabilities of informed trading, etc can help in designing complex yet powerful intensity functions. The relevant code for this work can be found at `https://github.com/RishiDarkDevil/Hawkes-Process-Price-Model`.