विद्याविनियोगाद्विकासः

Market Microstructures

Hawkes-process modeling for Self-Exciting Market Dynamics Study

Submitted to:

Instructor: Prof. Anirban Banerjee

Submitted by:

Mohammed Irfan T

PGPGC202400142

Table of Contents

Executive Summary

This project conducts an empirical study on the microstructure dynamics and endogeneity of high-frequency order flow in the National Stock Exchange (NSE) cash market. Focusing on a highly liquid equity (INFY) over a ten-day period in August 2019, the study applies Bivariate Hawkes Processes to model the self-excitation of buyer- and seller-initiated trades. The primary objective is to capture *endogenous clustering* of order flow: market activity leads to more activity (self-exciting). This ties directly to sequential trade models and information arrival theory in market microstructure, where clustered order flow can be interpreted as information cascades or algorithmic execution waves.

The methodology in this report contrasts two distinct model specifications: 1) a standard time-dependent counting process, Bivariate Hawkes Processes not accounting for volume of trade and 2) a Marked Bivariate Hawkes Process that incorporates trade volume as a multiplier in the excitation kernel. The analysis yields interesting insights. The market exhibits definitive evidence for clustering or endogeneity, implying that order flow at the microsecond scale is driven almost exclusively by feedback loops rather than external information. The findings ultimately characterize the asset as a highly reactive, machine-driven ecosystem where algorithmic order splitting is the dominant driver of clustering

An algo v/s non algo analysis done in this works shows us that while Algorithmic flows dominate market share (~63%) and execution speed, the 'Non-Algo' segment exhibits paradoxically higher instability and occasional microsecond-level reaction times. This suggests the presence of aggressive proprietary strategies within the non-institutional order flow, contradicting the traditional view of manual trading stability.

Introduction

Context

In the era of High-Frequency Trading (HFT), order flow exhibits distinct non-Markovian properties: trades do not arrive independently; rather, they arrive in bursts. A large buy order often leads to a series of subsequent buy orders (herding) or immediate sell orders (inventory rebalancing). To capture these dynamics, traditional Poisson processes are insufficient as they assume independent event arrivals. This project employs the **Hawkes Process**, a self-exciting point process, to model the arrival times of buyer-initiated and seller-initiated trades for a highly liquid equity (INFY) on the National Stock Exchange (NSE).

The Hawkes Process

The Hawkes process is used to model "self-exciting" events, where the occurrence of an event increases the probability of future events. While originally applied to earthquake aftershocks, it has become the gold standard in quantitative finance for modeling **order flow clustering**.

Mathematically, the Hawkes process is defined by its conditional intensity function $\lambda(t)$, which represents the probability of an event occurring in the next instant dt given the history of all previous events $\mathcal{H}_t$. Event, in our case would be buyer-initiated trade or seller-initiated trade.

$$\lambda(t) = \mu + \sum_{t_i<t} \phi(t - t_i)$$

This equation decomposes market activity into two distinct components:

1. **Exogenous Intensity (μ):** The baseline rate of events driven by external news, macro data, or fundamental value changes. This represents the true information flow.
2. **Endogenous Intensity (∑ ϕ)**: The internal feedback loop. Each past event t_i adds a temporary spike to the intensity, which decays over time according to a kernel $\phi(.)$

Why NSE?

In the context of an Order Driven Market like the NSE, Hawkes process modelling allows us to quantify **Market Reflexivity,** the degree to which the market is reacting to itself rather than to new information.

While majority of high-frequency literature focuses on developed markets (NYSE, NASDAQ, LSE), the National Stock Exchange of India (NSE) offers a unique case for microstructure analysis. As one of the world's largest exchanges by trade volume, the NSE is a pure Limit Order Book (LOB) market characterized by high retail participation along with sophisticated HFTs. The adoption of co-location servers and algorithmic trading in India has altered the latency landscape, effectively compressing the "memory" of the market from seconds to microseconds. Analyzing NSE data allows us to test whether the "endogeneity" observed in mature markets studied in the literature holds true in an emerging market structure, where tick sizes and liquidity replenishment rates differ.

Literature Review

The application of point processes to finance has a rich academic lineage:

- **Bowsher (2007)** was among the first to apply Hawkes processes to financial data, modelling the joint arrival of trades and mid-price changes.
- **Large (2007)** extended this to measure market resiliency, showing how the LOB replenishes liquidity after a trade consumes it.
- **Filimonov and Sornette (2012)** provided a seminal contribution by introducing the "Reflexivity Index" (or Branching Ratio).
- **Hardiman et al. (2013)** showed that financial markets often operate with branching ratios close to 1, implying strong endogeneity and raising the possibility of near **self-organized critical systems**.
- **Rambaldi et al. (2017)** demonstrated that incorporating trade size through marked Hawkes processes significantly improves modelling of order-flow dynamics, as trade volume carries additional informational content.

Scope of Empirical Analysis

Using tick-by-tick data for **Infosys (INFY)** from August 2019, I aim to perform a rigorous empirical analysis using a **Bivariate Hawkes Process** (modelling the interaction between Buy and Sell orders).

Specifically, this project will:

1. **Calibrate the Model:** Estimate the baseline intensities (**μ**), excitation parameters (**α**)**,** and decay rates (**β**) using Maximum Likelihood Estimation (MLE) optimized for large datasets.
2. **Quantify Endogeneity:** Calculate the **Branching Ratio (η)** and **Spectral Radius (ρ)** to determine the stability of the order book. A spectral radius close to 1 indicates a market in a "critical" state, close to zero exogeneity.

3. **Compare Model Specifications:** Contrast a standard (unmarked) time-dependent model against a **Volume-Marked Hawkes Process**. This comparison is critical to understanding whether market fragility is driven by the *frequency* of trades or the *volume* of liquidity consumed.
4. **Microstructure Interpretation:** Interpret the estimated kernel parameters to detect phenomena such as **Order Splitting** (diagonal dominance in the excitation matrix) vs. **Mean Reversion** (cross-excitation).
5. **Agent Segmentation:** Decompose order flow into Algorithmic and Non-Algorithmic components to test the Algo v/s. Non Algo hypothesis, specifically comparing the latency (β) and endogeneity (ρ) of institutional algorithms versus other market participants.

Data Preparation

The procedures employed ensure a high-fidelity reconstruction of the Limit Order Book (LOB) dynamics (not fully reconstructing the LOB, but extracting only the required details), specifically focusing on the rigorous classification of trade aggressors.

Data Source

The analysis is based on high-frequency tick-level data sourced from the National Stock Exchange of India (NSE). The raw data consists of two primary separate streams:

- Order Tick Data: Contains every order entry, modification, and cancellation.
- Trade Tick Data: Contains every executed trade

Scope and Duration

The study focuses on Infosys Ltd. (INFY), utilizing data from the Cash Segment, Series EQ. The dataset spans 10 trading days from August 13, 2019, to August 27, 2019 (excluding weekends and holidays). The specific dates processed are:

- Year: 2019
- Month: August
- Days: 13, 14, 16, 19, 20, 21, 22, 23, 26, 27.

Data Processing

The data preparation pipeline comprised three distinct stages:

1. Extraction
2. Aggressor Classification, &
3. Consolidation.

Extraction

A custom optimized data loader was developed to stream the large raw daily files. The following filtering criteria were applied during extraction:

- Segment: ‘CASH’
- Symbol: ‘INFY’
- Series: ‘EQ’

For each day, the raw fixed-width files were parsed into structured CSV formats. The schemas extracted are as follows:

Order Data Fields: record_type, segment, order_number, timestamp, side, activity_type (ENTRY/MODIFY/CANCEL), symbol, series, volume, limit_price, trigger_price, algo_indicator, client_type.

Trade Data Fields: trade_number, timestamp, trade_price, volume, buy_order_number, sell_order_number, buy_algo, sell_algo

Aggressor Classification

A critical step in OFI modeling is identifying the "aggressor", the side (buy or sell) that initiated the trade by crossing the spread. Since the raw trade data does not explicitly flag the aggressor, it was derived by reconstructing the order matching timeline. For every trade, the corresponding buy order number and sell order number were mapped to their original timestamp from the Order Data.

Let T_{buy} be the entry timestamp of the buy order and T_{sell} be the entry timestamp of the sell order. The aggressor side ϵ is determined as:

$$\epsilon = \{+1 \text{ (Buyer Initiated) if } (T_{buy} > T_{sell})$$

$$-1 \text{ (Seller Initiated) if } (T_{sell} > T_{buy})$$

Logic: In a Price-Time priority matching engine (like NSE), the passive order must already be resting in the book. The active order (aggressor) arrives later and matches immediately. Thus, the order with the later entry timestamp is the aggressor.

Edge Cases: If the sell order was missing from the records (e.g., IOC or market order nuances), the buy order was assumed successfully resting, implying the missing sell order was the aggressor ($\epsilon = -1$). Conversely, if the buy order was missing, the trade was classified as Buyer Initiated ($\epsilon = +1$)

Participant Segmentation

To analyze agent heterogeneity, the order flow was split into two different streams based on the Aggressor's identity. Using the NSE's algo_flag field, trades were classified as:

- Machine (Algo): Flags 0 (Algo) or 2 (Algo via Smart Order Router)
- Human (Non-Algo): Flags 1 (Non-Algo) or 3 (Non-Algo via SOR). This binary classification allows for the independent calibration of Hawkes processes for institutional and non-institutional flow

Consolidation

The processed daily trade files, now merged with determined aggressor side, buy entry timestamp, and sell entry timestamp, were concatenated into a single unified data frame for optimization and analysis.

Model Calibration and Optimization Procedure

To move from the processed trade file to the optimization exercise and interpretation of the microstructure signals, we need to run the Bivariate Hawkes Process maximizing Maximum Likelihood Estimation (MLE). Unlike standard approaches that rely on black-box libraries (such as tick or hawkeslib), this study utilizes a **self-built, high-performance solver** developed in Python. This approach **provided total control over the optimization**, allowing us to implement specific constraints required for high-frequency data, such as microsecond-level decay bounds and volume-weighted excitation kernels.

The Optimization Objective

The goal of the calibration is to find the specific set of parameters μ, α, and β that maximizes the probability that the observed sequence of trades (Buy and Sell) occurred exactly as recorded. We achieved this by minimizing the negative Log-Likelihood Function $\mathcal{L}$. The likelihood tries to find the parameters that make the intensity high when trades happen and low when trades don't happen.

For a bivariate process, the objective function is defined as:

$$\ln \mathcal{L}\,(\theta) = \sum_{i=1}^{N} \quad \ln \lambda_{m_i}\,(t_i) - \sum_{m=1}^{M} \int_0^T \lambda_m(t)\ dt$$

Where:

$$\lambda = \mu + \sum_{t_k<t} \quad \alpha * \ exp\ (-\beta(t - t_k))$$

λ (The Intensity): This is the Conditional Intensity Function. It represents the instantaneous probability (or "excitement level") of a trade happening *right now*

μ : Base level of intensity for an event type (buyer/seller-initiated trade)

α (jump): When a trade happens, λ is increased by a factor controlled by α

β (decay): After the jump, the effect of earlier trades on λ decays down. β controls how fast it slides.

N (Total Events) = Total number of trades in a day

M (Dimensions): This is the number of event types modelled. In this case there are 2 event types, where m = 1 implies buyer-initiated trades and m = 2 implies a seller-initiated trade.

The first term of the equation can be thought of as the reward when high intensity collides with actual trade occurrence. It looks at the exact timestamps where trades actually happened (t_i). The question we are trying to optimize for is "Did the model returned high intensity (λ) right at the moment a trade occurred?" That is, if a trade happened at 10:00:01, we want λ to be high at 10:00:01. The higher the λ at that moment, the higher the Likelihood score.

The second term can be thought of as the punishment for predicting trades that did not happen. This calculates the area under the curve of the intensity function over the whole day. This represents the total number of trades the model expected to happen. Without this term, the model could force λ to be huge everywhere. This term penalizes the model for predicting trades that did not happen.

For the Marked Model, this objective was modified to include the volume mark V_i scaling the excitation term by a factor of $log(1 + V_i)$

Custom Solver Implementation

Given the computational complexity of calculating the likelihood for millions of trades per day, a naive implementation would scale quadratically $O(N^2)$, rendering it infeasible. To overcome this, a recursive O(N) algorithm was implemented. Instead of summing over the entire history for every new trade, the solver maintains a state variable R(t) that tracks the accumulated excitation. As the solver moves from trade t_{i-1} to t_i, the previous excitation is decayed by $e^{-\beta(t_i - t_{i-1})}$ and the new impulse is added. Key Technical Features:

- Solver Engine: Utilized scipy.optimize.minimize with the L-BFGS-B algorithm (Limited-memory Broyden–Fletcher–Goldfarb–Shanno). This quasi-Newton method is ideal for highdimensional optimization with bounds, ensuring that intensities and rates remain strictly positive.

- Speed Optimization: The likelihood function was optimized using vectorized numpy operations and Just-In-Time (JIT) compilation concepts to handle the recursive loop efficiently. This allowed the model to process a full trading day (approx. 50,000 to 100,000 events) under 5 minutes.
- HFT-Specific Bounds: Standard libraries often restrict the decay parameter β to prevent instability. However, market microstructure operates at microsecond speeds. The custom solver was designed with relaxed upper bounds (allowing β up to 100,000), enabling the model to capture very short-term memory.

Variable definition & Economic Implementation

The optimization process generates a vector of parameters for each trading day, post the bivariate optimization process. Below is the dictionary of variables presented in the results, along with their interpretation in the context of market microstructure.

Variable Name	Symbol	Definition	Market Implication
mu_buy / mu_sell	μ	Baseline intensity	Low value implies highly endogenous, High value implies information driven
beta	β	The speed at which the market forgets a trade shock.	High value means market reacts and recovers in microseconds
alpha_xy	α	The instantaneous jump in intensity of type x caused by an event of type y	Used to calculate branching ratios (
br_bb	Γ_{BB}	Branching Ratio: The expected number of Buy trades triggered by a single Buy trade	High value is evidence for herding or order splitting
br_bs	Γ_{BS}	Branching Ration: The expected number of Buy trades triggered by a single Sell trade	High value is evidence for mean reversion
avg_mark	κ	The mean of log(1+Volume)	Measures the average magnitude of volume impact. Used to scale the branching ratio in the Marked model
spectral_radius	ρ	Stability Metric. The largest eigenvalue of the branching matrix.	<1 : Stable Market or the shock dies out

Generation of Results

The optimization routine was executed iteratively across the 10-day dataset (August 13–27, 2019). For each day, the solver:

- Ingested the clean, signed order flow (and volumes for the Marked model)

- Applied a Stable Sort to ensure deterministic ordering of trades occurring at the same microsecond
- Performed the MLE minimization to extract the optimal parameter set
- Validated the solution by checking the Stationarity Condition

Analysis of Results

This section presents the empirical findings from the calibration of the Bivariate Hawkes Process on the INFY order book for the ten-day period in August 2019. The analysis is divided into two parts: first, the baseline "Tick-Count" model (Unmarked), which treats every trade equally; and second, the "Volume-Marked" model, which weights excitation by trade size.

Unmarked Bivariate Model: The "Tick" View

The basic question being answered in this analysis is *"How does the occurrence of a trade trigger future trades"*. The parameters after optimization, for all 10 trading days is shown below:

	mu_buy	mu_sell	alpha_bb	alpha_bs	alpha_sb	alpha_ss	beta
2019-08-13	0.0	0.000016	11110.233936	4328.096229	3885.044399	9725.681382	24563.963993
2019-08-14	0.000011	0.0	48099.381852	31626.902075	33895.498939	45455.45169	99999.813684
2019-08-16	0.000068	0.0	4372.033916	356.632035	142.494795	2008.722547	6856.359803
2019-08-19	0.0	0.000078	28664.668389	23017.421771	33093.052836	23789.058743	88075.556987
2019-08-20	0.0	0.0	25399.988192	25188.971964	369.792795	25445.02027	51008.804132
2019-08-21	0.0	0.000016	48383.310305	6742.665147	6984.647903	49660.39007	100000.0
2019-08-22	0.0	0.000095	44952.360015	31921.193859	27910.666324	45055.379387	83601.905083
2019-08-23	0.000376	0.0	25019.183007	2787.515184	2405.62688	5889.527255	50521.655286
2019-08-26	0.000079	0.0	39366.477002	11141.328246	28240.157235	39356.352995	80812.365558
2019-08-27	0.0	0.000024	6510.265654	5275.861555	714.876389	6021.775821	13669.339304

	br_bb	br_bs	br_sb	br_ss	spectral_radius	total_branching	log_likelihood	is_stationary
2019-08-13	0.452298	0.176197	0.15816	0.395933	0.593413	1.182588	1470278.028401	True
2019-08-14	0.480995	0.31627	0.338956	0.454555	0.795458	1.590775	1550959.230572	True
2019-08-16	0.637661	0.052015	0.020783	0.292972	0.640769	1.003431	994694.178846	True
2019-08-19	0.325455	0.261337	0.375735	0.270098	0.612355	1.232626	1042637.65104	True
2019-08-20	0.497953	0.493816	0.00725	0.498836	0.558229	1.497855	1287516.942039	True
2019-08-21	0.483833	0.067427	0.069846	0.496604	0.559141	1.11771	1552346.165708	True
2019-08-22	0.537695	0.381824	0.333852	0.538928	0.895345	1.792299	1352542.297152	True
2019-08-23	0.495217	0.055175	0.047616	0.116574	0.502033	0.714582	1272864.913124	True
2019-08-26	0.487134	0.137867	0.349453	0.487009	0.706566	1.461463	1401643.833905	True
2019-08-27	0.476268	0.385963	0.052298	0.440532	0.601593	1.35506	2123413.085598	True

Table 1: Unmarked (Tick) Bivariate Model Optimization Results

Key observations from the results are:

- System Stability ($\rho < 1$): Across all ten trading days, the estimated Spectral Radius (ρ) remains strictly below 1.0, ranging from 0.50 to 0.80. This confirms that the INFY order book is stationary: despite high-frequency bursting, the market possesses sufficient "damping" to prevent infinite cascading volatility.

- Timescale of Reaction (β): The estimated decay parameters (β) are consistently large, ranging from 20,000 to 100,000. In temporal terms, this corresponds to a "half-life" of information ($\ln(2)/\beta$) between 7 and 35 microseconds. This provides definitive evidence that the clustering mechanism possibly driven by High-Frequency Trading (HFT) algorithms. Human reaction times are in the range of 200 milliseconds; the observed dynamics are nearly 5,000 times faster, reflecting the speed of co-located, low latency execution engines
- μ close zero: On almost all days, mu_buy and mu_sell are close to 0. Which implies the otder flow is predominantly endogenous. At the microsecond scale, trades are rarely driven by new "news"; they are almost always reactions to previous trades
- Anomalies and Bound Constraints: On August 21st , the β parameter hit the upper optimization bound of 100,000. This suggests that on 21st the market's memory was effectively instantaneous, likely driven by intense algorithms.
- Analyzing the branching matrix (not scaled by β): The branching matrix tells us "If one event of type j happens, how many events of type i does it trigger?" Taking the example of branching matrix on 13th August,

$$\begin{pmatrix} buy-buy & sell-buy \\ buy-sell & sell-sell \end{pmatrix} = \begin{pmatrix} 11{,}110.24 & 3885.04 \\ 4328.10 & 9725.68 \end{pmatrix}$$

  We can see that the coefficients of buy-buy and sell-sell (diagonal elements) are more than double the non-diagonal elements. This is evidence of Order Splitting. Large institutional orders are broken into child orders that execute sequentially, creating "runs" of Buys. This pattern is seen throughout the 10-days of analysis, except 19th August, where sell-buy is the highest coefficient. Additionally, the non-diagonal elements are non-zero. A meaningful conclusion we can draw from this is a Buy does eventually trigger Sells (and vice versa), likely due to market makers stepping in to rebalance their inventory or capture the spread after a run.

Marked Bivariate Model: The "Volume-marked" View

The Marked model incorporates trade volume, weighting the kernel by log(1 + V). The parameters after optimization are as below:

	mu_buy	mu_sell	alpha_bb	alpha_bs	alpha_sb	alpha_ss	beta	avg_mark
2019-08-13	0.0	0.00005	1413.224351	1351.644465	484.358717	1569.618928	10376.059644	2.716042
2019-08-14	0.000058	0.0	2972.043103	603.450777	647.627549	3214.557979	10530.310853	2.655542
2019-08-16	0.001694	0.0	1399.760427	289.769062	643.928139	1500.988531	4998.900858	2.600939
2019-08-19	0.0	0.001814	2503.635706	751.308251	1012.084692	2724.352546	9037.600882	2.480952
2019-08-20	0.030899	0.0	25357.324177	10285.932356	4820.213991	24472.820543	100000.0	2.903157
2019-08-21	0.0	0.009954	4146.171295	1055.775858	1277.848426	4887.623857	15304.862688	2.497078
2019-08-22	0.0	0.001064	2359.33165	469.465901	935.59028	2714.553271	7941.134274	2.422804
2019-08-23	0.00613	0.0	1011.323316	131.222265	277.351332	1193.114915	3745.769801	2.571068
2019-08-26	0.002368	0.0	8029.707875	2193.063925	1648.505278	9464.41772	30228.878863	2.356927
2019-08-27	0.0	0.000305	1535.116308	196.761947	207.427762	2566.615916	7606.161419	2.772518

	br_bb	br_bs	br_sb	br_ss	spectral_radius	total_branching	log_likelihood	is_stationary
2019-08-13	0.369926	0.353807	0.126786	0.410864	0.603178	1.261383	1351100.134825	True
2019-08-14	0.749492	0.152179	0.163319	0.81065	0.94066	1.87564	1324903.832932	True
2019-08-16	0.728298	0.150767	0.335037	0.780968	0.980921	1.99507	989368.7902	True
2019-08-19	0.687284	0.206245	0.277832	0.747874	0.958866	1.919235	882652.692502	True
2019-08-20	0.736163	0.298617	0.139938	0.710484	0.928147	1.885202	1398866.893976	True
2019-08-21	0.676472	0.172256	0.208488	0.797444	0.935885	1.854661	1356264.773498	True
2019-08-22	0.719821	0.143232	0.285444	0.828198	0.983345	1.976696	1128120.718765	True
2019-08-23	0.694165	0.09007	0.190372	0.818945	0.901604	1.793552	1058740.036855	True
2019-08-26	0.626071	0.170992	0.128533	0.737935	0.840453	1.663531	1304165.228145	True
2019-08-27	0.559564	0.071722	0.075609	0.935556	0.949464	1.642451	2028976.670924	True

Table 2: Volume Marked Bivariate Model Optimization Results

Key observations from the analysis are:

- Spectral Radius Shift: The most important finding is that there is an increase in the Spectral Radius when volume is considered. While the Unmarked ρ is close to 0.9 (almost unstable).
    - Implication: The market is far more unstable than what the tick-count model reveals. A single large order can trigger a disproportionately long and volatile chain reaction.
- Extended Memory: On average, the half-life of a volume-weighted shock was roughly 3x longer than that of a simple tick shock. This is evidence for large trades depleting liquidity at a specific level which leads to market taking physical time to replenish, influencing trader behavior for much longer and hence larger half-life.

Heterogeneity Analysis – Algo v/s Non Algo

By fitting separate Bivariate Hawkes models to the Algo and Non-Algo order streams, we observe a market structure that defies the belief that non-algo's are slow and exogeneous.

date	count_algo	count_non-algo	algo_beta	algo_half_life_us	algo_spectral_radius	algo_br_diagonal	non-algo_beta	non-algo_half_life_us	non-algo_spectral_radius	non-algo_br_diagonal	speed_ratio
2019-08-13	103539	57850	4318.036304	160.523704	0.509442	0.824103	34369.396397	20.167569	0.988593	0.996020	0.125636
2019-08-14	103708	50748	58094.785873	11.931315	0.502562	0.997605	32601.421763	21.261256	0.628434	0.926066	1.781971
2019-08-16	80433	44630	12005.436898	57.736106	0.589020	0.880280	50860.985808	13.628269	0.583091	1.003683	0.236044
2019-08-19	68209	39093	44430.287019	15.600781	0.876762	0.909859	34933.424117	19.841948	0.983403	0.994661	1.271856
2019-08-20	83063	57437	62997.960467	11.002692	0.791374	0.933342	57840.201953	11.983831	0.919616	1.187333	1.089173
2019-08-21	98704	55812	22228.903251	31.182248	0.628436	0.886825	7872.205751	88.049932	0.643413	0.884400	2.823720
2019-08-22	89120	47039	50636.711069	13.688630	0.532725	0.998439	14612.204620	47.436181	0.586860	0.837943	3.465371
2019-08-23	88442	53205	100000.000000	6.931472	0.774694	1.000000	59523.319427	11.644969	0.570498	0.998274	1.680014
2019-08-26	85167	56162	33819.104960	20.495728	0.711316	1.189391	24034.420474	28.839771	0.636617	0.913826	1.407111
2019-08-27	151176	90436	51039.810411	13.580520	0.541800	0.997391	4475.424125	154.878546	0.592257	0.875988	11.404463

Table 3: Segmentation results summary of Algo v/s Non Algo Trades

Algorithmic Dominance: As expected, the Algo segment accounts for ~63.3% of all aggressive (trade initiation side is the algo side) trades and exhibits a higher average reaction speed or decay rate(β_{algo} $close\ to$ 44,000 and $\beta_{non-algo}$ close to 32,000)

Contrary to the hypothesis that institutional algorithms drive volatility, the "Non-Algo" segment consistently displayed higher instability. On 7 out of 10 days, the Non-Algo spectral radius ($\rho \approx 0.71$) exceeded that of the Algo segment $\rho \approx 0.65$ reaching critical levels $\rho \approx 0.98$ on August 13th .

We can also note some latency anomalies. On multiple trading days, the Non-Algo beta exceeded 34,000 (a reaction time of $\approx 20\ \mu s$), which is physically impossible for manual traders. These results points towards the presence of momentum driven on

date	mu_buy	mu_sell	alpha_bb	alpha_bs	alpha_sb	alpha_ss	beta	br_bb	br_bs	br_sb	br_ss	spectral_radius	total_branching	log_likelihood	is_stationary
2019-08-13	1.000000e-10	1.037190e-03	1864.444766	251.825442	673.452421	1694.062262	4318.036304	0.431781	0.058319	0.155963	0.392322	0.509442	1.038385	8.188353e+05	True
2019-08-14	1.000000e-10	1.293475e-06	28979.303389	210.173193	226.950145	28976.321684	58094.785873	0.498828	0.003618	0.003907	0.498777	0.502562	1.005129	9.747677e+05	True
2019-08-16	8.558101e-04	1.000000e-10	5373.544047	1486.268442	2144.098795	5194.601425	12005.436898	0.447593	0.123800	0.178594	0.432687	0.589020	1.182674	6.910911e+05	True
2019-08-19	1.000000e-10	9.565340e-04	19861.982629	18698.519071	18779.288439	20563.323822	44430.287019	0.447037	0.420851	0.422669	0.462822	0.876762	1.753379	6.373506e+05	True
2019-08-20	1.000000e-10	9.840151e-02	27758.627539	19329.981323	21507.576886	31040.004994	62997.960467	0.440627	0.306835	0.341401	0.492714	0.791374	1.581578	8.021317e+05	True
2019-08-21	1.000000e-10	1.176339e-02	8927.421497	3987.190847	4026.000467	10785.715478	22228.903251	0.401613	0.179370	0.181116	0.485211	0.628436	1.247310	8.895245e+05	True
2019-08-22	5.390653e-05	1.000000e-10	25272.587700	114.485254	25142.382986	25285.072298	50636.711069	0.499096	0.002261	0.496525	0.499343	0.532725	1.497225	8.267375e+05	True
2019-08-23	1.000000e-10	1.842920e-01	50000.000000	24352.437001	30985.396114	50000.000000	100000.000000	0.500000	0.243524	0.309854	0.500000	0.774694	1.553378	8.849105e+05	True
2019-08-26	1.011721e-02	1.000000e-10	21791.425566	5169.138395	2463.632301	18432.728103	33819.104960	0.644353	0.152847	0.072847	0.545039	0.711316	1.415085	8.021493e+05	True
2019-08-27	1.000000e-10	1.402832e-07	25456.097320	192.612043	25129.653603	25450.557441	51039.810411	0.498750	0.003774	0.492354	0.498641	0.541800	1.493519	1.424706e+06	True

Table 4: Algo Trades – Bivariate (unmarked/tick) Hawkes Model Optimization results

date	mu_buy	mu_sell	alpha_bb	alpha_bs	alpha_sb	alpha_ss	beta	br_bb	br_bs	br_sb	br_ss	spectral_radius	total_branching	log_likelihood	is_stationary
2019-08-13	1.000000e-10	6.572363e-05	17169.798409	16880.308238	16841.619653	17062.805509	34369.396397	0.499566	0.491144	0.490018	0.496453	0.988593	1.977181	532219.520248	True
2019-08-14	5.944113e-05	1.000000e-10	16331.287116	5821.140954	4732.708245	13859.790106	32601.421763	0.500938	0.178555	0.145169	0.425128	0.628434	1.249790	472520.987598	True
2019-08-16	1.000000e-10	2.151539e-04	25506.010830	25086.971611	680.697894	25542.297905	50860.985808	0.501485	0.493246	0.013383	0.502198	0.583091	1.510312	418712.007891	True
2019-08-19	1.000000e-10	4.148919e-05	17418.204132	16987.224526	16973.017950	17328.707442	34933.424117	0.498611	0.486274	0.485868	0.496049	0.983403	1.966803	358673.443920	True
2019-08-20	2.034588e-05	1.000000e-10	34334.733198	12346.149727	28789.151167	34340.856560	57840.201953	0.593614	0.213453	0.497736	0.593720	0.919616	1.898522	550910.180933	True
2019-08-21	1.000000e-10	2.745550e-05	3337.060055	1330.395698	1870.320632	3625.121453	7872.205751	0.423904	0.168999	0.237585	0.460496	0.643413	1.290985	461517.932875	True
2019-08-22	1.000000e-10	6.236576e-05	6248.376667	3036.096738	1977.010384	5995.812695	14612.204620	0.427614	0.207778	0.135299	0.410329	0.586860	1.181019	411708.990939	True
2019-08-23	4.974984e-05	1.000000e-10	29709.110330	614.393685	29366.224793	29711.497738	59523.319427	0.499117	0.010322	0.493357	0.499157	0.570498	1.501953	502827.982361	True
2019-08-26	1.516089e-04	1.000000e-10	10977.984900	3768.766666	4949.741461	10985.297229	24034.420474	0.456761	0.156807	0.205944	0.457065	0.636617	1.276577	511116.014326	True
2019-08-27	1.000000e-10	8.799482e-05	1931.613153	941.439769	505.421924	1988.803536	4475.424125	0.431604	0.210358	0.112933	0.444383	0.592257	1.199278	715764.186137	True

Table 4: Non-Algo Trades – Bivariate (unmarked/tick) Hawkes Model Optimization results

Visualization

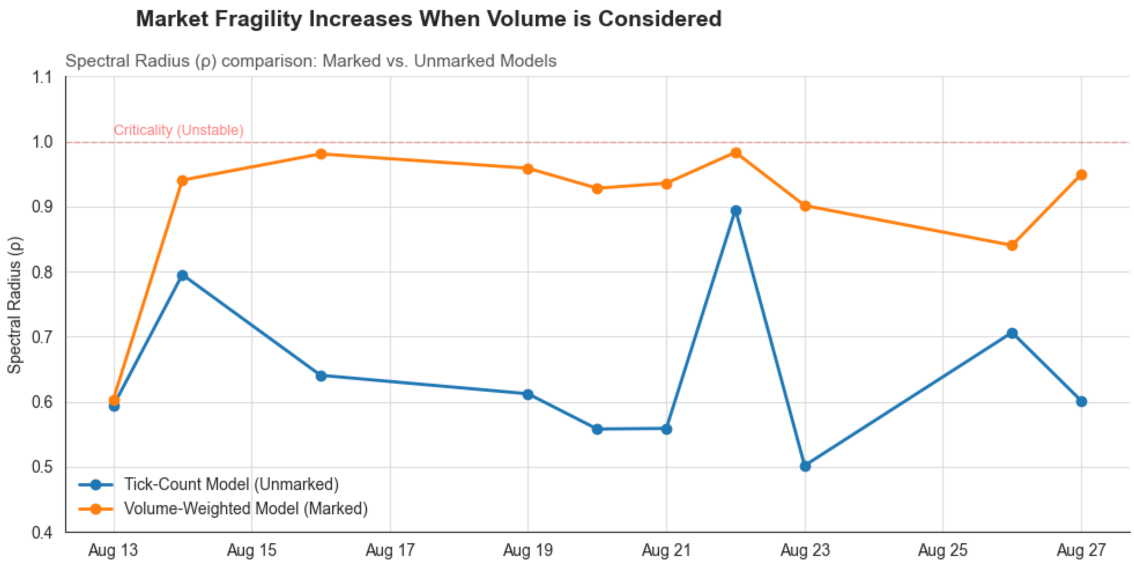

Figure 1: Marked v/s Unmarked Model Stability (Spectral Radius)

This above figure illustrates the hidden risk in simple tick analysis. The Orange line (Marked Model) consistently tracks higher than the blue line (Tick Model), visually demonstrating that taking volume into consideration, brings out the hidden facts about stability. The proximity of the orange line to the red "Criticality" threshold indicates a market operating on the edge of instability.

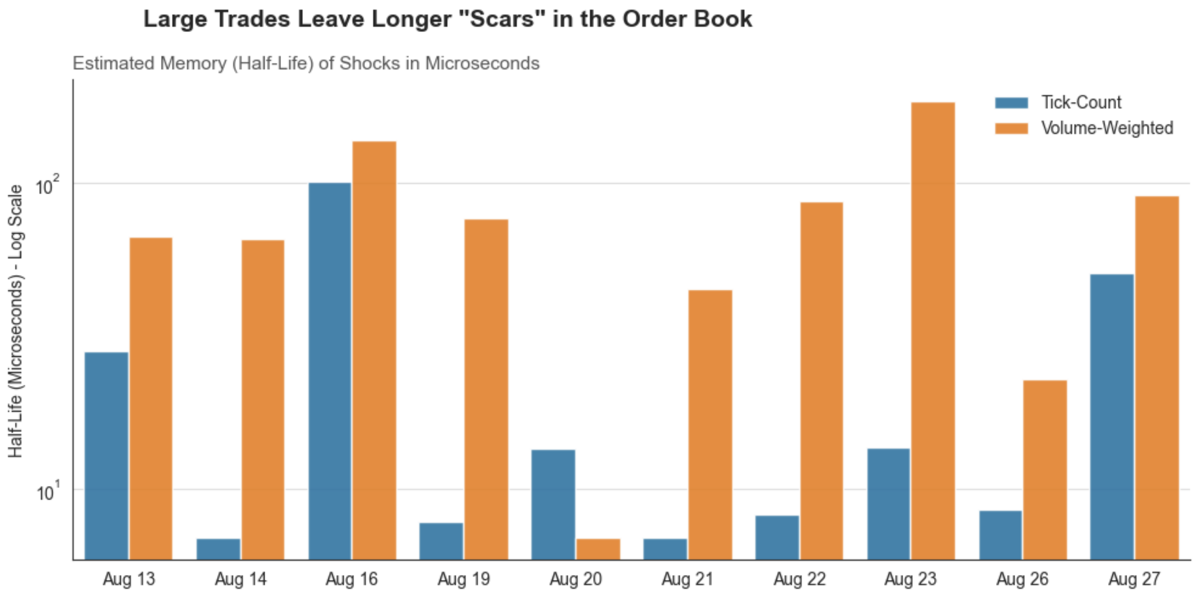

Figure 2: Marked v/s Unmarked Model - Memory (Half-life (= $ln(2)/\beta$))

The logarithmic bar chart compares the Half-Life of shock from a single trade. The significantly taller orange bars show that the market remembers volume. A large trade does not just vanish; it alters the probability of future trade for a significantly longer window than small trade.

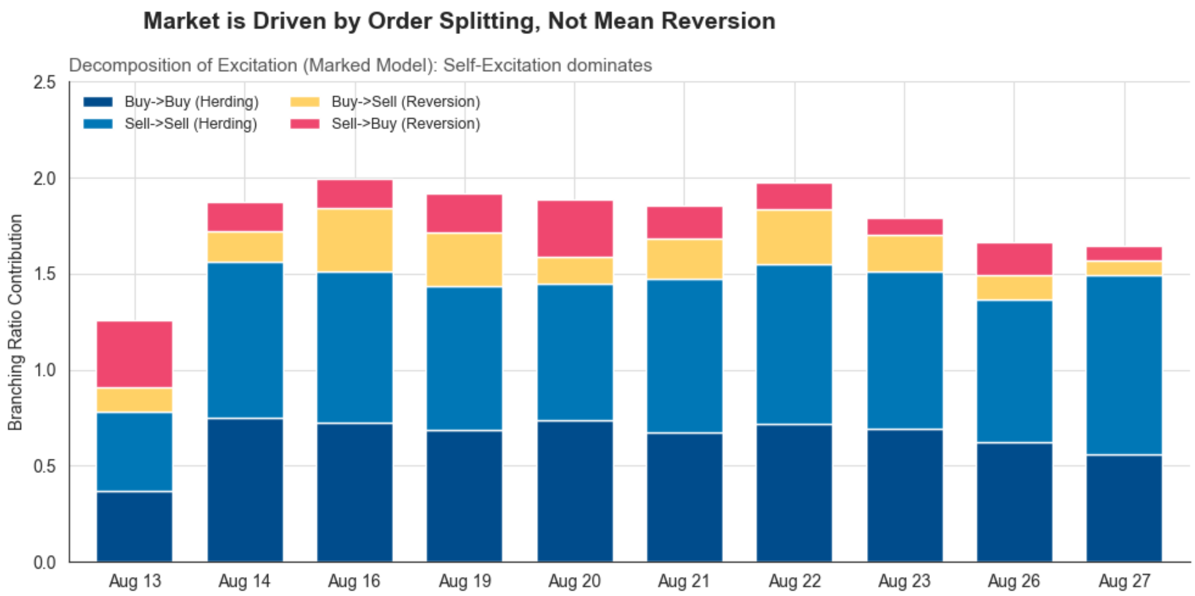

Figure 3: Decomposition of Herding/Order Splitting effects (the two blue bars) and Mean Reversion (yellow and red bars)

Conclusion

This project establishes that high-frequency order flow on the National Stock Exchange for the security of choice, INFY, shows a high degree self-exciting characteristics. By applying both unmarked and volume-marked Bivariate Hawkes Processes, the analysis shows us that the market's memory is longer and more unstable than a simple trade-count model would suggest. The project also finds that large-volume trades leave longer impact in the order book, effectively changing trade probabilities for significantly longer time. Furthermore, the segmentation of Algorithmic vs. Non-Algorithmic flows finds a counter-intuitive result: while algorithms dominate speed and volume, the "Non-Algo" segment exhibits higher structural instability and occasional microsecond-level latency. Ultimately, the study gives enough evidence that the NSE operates in a near-critical state where feedback loops and self-excitation, rather than external news, drive most of the micro-scale level behaviors.

References

- Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. Journal of Econometrics, 141(2), 876-912.
- Large, J. (2007). Measuring the resiliency of an electronic limit-order book. Journal of Financial Markets, 10(1), 1-25.
- Filimonov, V., & Sornette, D. (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. Physical Review E, 85(5), 056108.
- Hardiman, S. J., Bercot, N., & Bouchaud, J. P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. The European Physical Journal B, 86, 442.
- Rambaldi, M., Bacry, E., & Lillo, F. (2017). The role of volume in order book dynamics: a multivariate Hawkes process analysis. Quantitative Finance, 17(7), 999-1020.