Order Book Filtration and Directional Signal Extraction at High Frequency

Aditya Nittur Anantha, Shashi Jain, Prithwish Maiti

Abstract

With the advent of electronic capital markets and algorithmic trading agents, the number of events in tick-by-tick market data has exploded. A large fraction of these orders is transient. Their ephemeral character degrades the informativeness of directional alphas derived from the limit order book (LOB) state. We investigate whether directional signals such as order book imbalance (OBI) can be improved by structurally filtering high-frequency LOB data. Three real-time, observable filtration schemes: based on order lifetime, update count, and inter-update delay. These are used to recompute OBI on structurally filtered event streams. To assess the effect of filtration, we implement a three-layer diagnostic framework: contemporaneous correlation with returns, explanatory power under discretized regime counts, and causal coherence via Hawkes excitation norms. Empirical results show that structural filtration improves directional signal clarity in correlation and regime-based metrics, but leads to only limited gains in causal excitation strength. In contrast, OBI computed using trade events exhibits stronger causal alignment with future price movements. These findings highlight the importance of differentiating between associative and causal diagnostics when designing high-frequency directional signals.

1 Introduction

Modern electronic markets generate thousands of order book updates every second. These reflect the activity of algorithmic traders, market makers, and latency-sensitive agents. While the abundance of microstructural data offers fine-grained visibility, it does not guarantee signal quality.

The limit order book (LOB) records all outstanding buy and sell orders at discrete price levels. However, this representation is not a direct window into latent trading interest. It reflects a combination of behaviors, including genuine limit order submission, quote replenishment, liquidity sampling, and high-frequency cancellation. These actions reshape the LOB without necessarily conveying directional intent. As a result, high-frequency indicators based on raw order flow are often contaminated by transient activity that masks meaningful signals.

Among the tools developed to extract directional signals from the limit order book, the *Order Book Imbalance (OBI)* stands out for its simplicity and operational tractability. Defined as the signed difference between aggregated bid and ask-side order flow, OBI reflects short-term pressure exerted by market participants. It serves as a concise proxy for microstructural disequilibrium [1, 2, 3], with widespread adoption driven by its real-time computability and intuitive appeal.

However, OBI is highly sensitive to transient distortions. In modern markets, OBI computed over the full stream of order events is vulnerable to flickering liquidity, orders rapidly submitted and canceled without execution intent. These patterns, driven by latency arbitrage, passive replenishment, or spoofing strategies [4, 5], introduce noise into the OBI signal. When left unfiltered, OBI conflates genuine imbalance with short-lived manipulation, weakening its directional informativeness and diminishing its relevance for short-term return analysis.

1.1 High-Frequency Trading and LOB Signal Noise

High-frequency trading (HFT) has fundamentally altered limit order book (LOB) dynamics by accelerating the speed of order submissions and cancellations. Competition for speed has led to an HFT

"arms race" where traders engage in costly technology upgrades to gain microsecond advantages [6]. One consequence of this arms race is the prevalence of fleeting orders or flickering quotes: orders that are placed and then canceled almost immediately, often within milliseconds. Empirical studies find that such fleeting orders are widespread in modern electronic markets and are closely associated with HFT activity [5]. For example, Hasbrouck and Saar document that a significant fraction of orders in the millisecond environment are canceled very quickly ("fleeting"), indicating that many displayed quotes are not intended to be executed [4]. These rapid-fire order updates can momentarily improve displayed liquidity or test market depth, but they inject substantial noise into the order flow. Indeed, nearly all limit orders ultimately get canceled rather than executed – recent research shows that the vast majority of submitted orders are withdrawn before trading, as traders update their quotes in response to changing information [7]. This environment complicates the extraction of reliable trading signals, since the LOB is cluttered with ephemeral information that may not reflect genuine supply and demand. As a result, researchers and practitioners have considered various filtering techniques to distinguish meaningful order book imbalances from mere noise. For instance, one common approach is to ignore or down-weight quotes that persist for only a few milliseconds, on the premise that such ultra-fast cancellations are strategic and not indicative of true interest. The motivation for filtered order book signals is to enhance signal quality by removing these microstructural artifacts, thereby focusing on more persistent order flow patterns. Market design proposals like frequent batch auctions have even been suggested to inherently reduce the noisy volatility caused by continuous-time racing and flickering quotes [6]. Overall, the rise of low-latency trading has improved certain liquidity measures (e.g., tighter spreads and greater depth) [4], but it has also introduced high-frequency microstructure noise. This duality has spurred interest in whether filtering the LOB data can better isolate the informative content (e.g,. true buying or selling pressure) from transient fluctuations.

1.2   Order Book Imbalance as a Predictive Signal

One of the key microstructural indicators for directional price pressure is the order book imbalance (OBI), also often termed order flow imbalance (OFI). OBI generally measures the difference between buy and sell interest in the limit order book, such as the volume (or number of orders) at bid prices versus ask prices. [3] formally introduced a definition of OFI based on detailed order book events, and demonstrated a striking linear relationship between order flow imbalance and short-term price changes. In their empirical analysis, price changes over short intervals were largely explained by the net order imbalance: when buy orders (market buys and limit orders at the bid) exceed sell orders (and cancellations) in aggregate, prices tend to rise, and vice versa. This linear price impact of imbalance had a slope inversely proportional to market depth, consistent with intuition that deeper order books absorb imbalance with less price movement [3]. The predictive power of imbalance has since been corroborated in various markets. For example, [1] find significant effects of daily order flow imbalance on returns in the German stock market. Their results show that days with excess buying (positive imbalance) generally see upward price pressure, and they also document that the strength of this effect varies with firm size and liquidity [1]. These studies cement the idea that OBI contains valuable information about the likely direction of price moves, making it a natural candidate for trading signals.

In practice, there is evidence that HFT firms and algorithmic traders utilize order book imbalance as part of their strategies. Though proprietary strategies are seldom disclosed, the correlation between OBI and future returns suggests informed traders monitor this metric [8]. Yagi et al. simulate an agent-based market in which one agent type engages in an OBI-driven high-frequency strategy, and they confirm that exploiting imbalance can yield a trading advantage [8]. In their simulation, an HFT that buys (sells) when the book is bid-imbalanced (ask-imbalanced) realizes profits, albeit with higher volatility – a finding that aligns with the notion that OBI-based strategies are profitable but can carry high risk in turbulent periods. The broader implication is that imbalance signals are indeed directional: a persistently buy-imbalanced book often precedes upward price movements, whereas a sell-imbalanced book presages downward moves. This principle underlies many short-term prediction models in the literature.

Recognizing the importance of imbalance, recent work has focused on extracting even more information from the limit order book beyond a simple aggregate imbalance. [2] propose a multi-horizon and

multi-level approach to OBI, dubbed "Deep Order Flow Imbalance." They construct features that capture imbalance across different layers of the book and across various time scales, and apply deep learning to combine these signals. The study finds that this enriched representation of order flow imbalance can improve the prediction of price movements at horizons ranging from intraday minutes to longer intervals [2]. In essence, while the original OBI measures were scalar summaries (often just considering the best bid/ask or cumulated top levels), the modern approach is to leverage the full depth and temporal evolution of the order book to extract latent signals. This reflects an evolution in signal construction: from simple static imbalances to complex, data-driven representations that capture the depth-wise and time-wise structure of order flow. The success of such models underscores that there is additional alpha to be mined by looking at how imbalance manifests at different scales, but it also raises the bar on data processing and noise reduction, since a multi-level view is even more susceptible to flickering at each price level if not filtered appropriately.

1.3 Filtered vs. Unfiltered Signals and Data Cleaning

The distinction between filtered and unfiltered OBI signals has become an important practical consideration in high-frequency data analysis. An unfiltered order book signal is computed directly from the raw data stream, counting every order placement, modification, cancellation, or trade that occurs. By contrast, a filtered signal applies criteria to ignore certain transient events or out-of-band quotes so as to capture a cleaner picture of supply and demand. The need for filtering arises from the realization that many order book updates are noise or strategy-induced artifacts. As noted, Ulze et al. find that flickering quotes (orders that flash in and out) can momentarily tighten spreads or give the illusion of depth, only to vanish before any trade occurs [5]. These flickers can cause abrupt swings in a naive imbalance measure computed tick-by-tick. Similarly, the extremely high cancellation rates documented by [7] imply that raw order flow data contain numerous cancellations that quickly negate earlier placements, effectively canceling out their impact. Including all such ephemeral updates in a signal may therefore add volatility without improving informational content.

To address this, researchers have explored filtering methods. A common approach is time-based filtering: for instance, define a quote as fleeting if it persists for less than a certain threshold (e.g. 50 milliseconds), and exclude those from imbalance calculations. This kind of filter was implicitly used in studies like Hasbrouck and Saar's analysis of low-latency trading, where they categorize fleeting orders as not contributing to genuine liquidity [4]. Another approach is event-based filtering, such as focusing only on imbalance changes caused by trades (executions) and large persistent orders, rather than every minor cancel/replace event. By doing so, one captures the net order flow that moves the price. In essence, filtered OBI metrics attempt to measure the structural imbalance, the imbalance that remains after fast tactical orders have been removed. While there is no universal filtering standard, empirical work suggests that filtering can substantially stabilize the imbalance signal and improve its correlation with price movements. For example, if one filters out flickering quotes, the resulting OBI time series tends to be smoother and may align more clearly with future price trends (this is often observed in proprietary trading analyses, even if not always formally published).

It should be noted that filtering is not without trade-offs. Aggressive filtering might throw away information, as some fleeting orders could be placed by informed traders testing the market. The academic literature has begun to grapple with these issues: recent studies highlight the importance of understanding HFT order add-and-cancel patterns when designing predictive signals. For instance, [6] argues that many rapid order updates are symptomatic of the race conditions in continuous markets, implying that a more coarse-time market structure (batches) could naturally filter out meaningless micro-events. Meanwhile, [8] and others illustrate that the effectiveness of an imbalance-based strategy can depend on market conditions, in highly volatile "flash crash" scenarios, even a filtered signal may fail, as market dynamics break down in extreme ways. Consequently, ongoing research often evaluates signal quality under different filtering schemes to ensure robustness. In summary, the consensus emerging is that unfiltered OBI signals reflect the noisy reality of high-frequency order flow, whereas filtered OBI signals aim to extract the enduring informational content by removing the shortest-lived, potentially spurious order book fluctuations. Careful filtration thus appears to enhance directional signal quality in many high-frequency trading applications, though the optimal filtering rules may vary by market and

strategy.

1.4 Point Process Models and Self-Exciting Order Flows

Beyond handcrafted signals, researchers have developed statistical models to capture the dynamics of order flow and imbalance in a more principled way. A prominent class of such models treats the sequence of order book events (limit orders, market orders, cancellations) as realizations of point processes (random event sequences in continuous time). Early work by [9] applied multivariate point process models to financial market events, using intensity functions (hazard rates) that could depend on the history of order arrivals. This intensity-based framework allowed for analyzing how the occurrence of one type of event (say, a trade) affects the short-term probability of another event (like a cancellation or a quote update). Building on this, a particularly influential model class in recent years has been Hawkes processes, introduced originally by [10]. Hawkes processes are self-exciting point processes, meaning each event increases the rate of future events for some period, thereby generating clustered event arrivals. This property maps well to order book dynamics, where trades and orders tend to cluster in bursts of activity.

In market microstructure, Hawkes processes have been extensively used to model order flows and even price changes. [11] provides a comprehensive overview of Hawkes process applications in finance, noting that these models can reproduce stylized facts such as heavy-tailed volume clustering and volatility clustering. For example, a Hawkes model can be fitted to the arrival times of market orders and limit orders; a significant self-excitation parameter would indicate that a flurry of buy orders tends to trigger further buy orders, capturing a feedback loop often associated with trend momentum or order anticipation. [12] use Hawkes processes to analyze the reflexivity of financial markets, essentially, how order flow can become self-reinforcing. They find that certain market events exhibit cascade effects consistent with a self-exciting process, which provides a statistical description of phenomena like herding or positive feedback in order flow. However, modeling high-frequency data with Hawkes processes is not without challenges. [13] caution that standard Hawkes model calibrations can misinterpret long-memory in order flows as "critical" self-excitation. They show that improper inference can falsely suggest that the system is on the verge of an unstable chain reaction, whereas in reality the persistence might be due to structural features or exogenous influences. This highlights the need for careful specification and calibration when applying point process models to LOB data.

Extensions of the basic Hawkes framework have also been proposed to incorporate more information. For instance, Rambaldi et al. develop a multivariate Hawkes process that includes order volume as part of the state, thus enriching the model beyond mere event timing [14]. Their analysis, focusing on how the size of orders influences future activity, suggests that volume plays a significant role in the excitation: a large trade can trigger a series of subsequent orders (potentially as other participants react to the large trade), a dynamic that a volume-aware Hawkes model can capture. As the dimensionality of these models grows (multiple event types, volume levels, etc.), traditional parametric estimation becomes complex. In response, researchers have turned to machine learning for help. [15] propose a neural network-based model for multi-dimensional nonlinear Hawkes processes. By using neural networks to flexibly parametrize the excitation kernel (the function that governs how one event influences the future rate of events), they can capture nonlinear dependencies and interactions between a wide array of order flow events. Such a model can, for example, learn if a burst of small buy orders followed by one large sell order creates a different excitation pattern than vice versa, nuances that would be hard to pre-specify in a classical model. These advanced point process models represent a shift from static signal indicators like OBI to *dynamic modeling* of imbalance: instead of outputting a single imbalance number, the model encodes the history of order flow and can forecast the likelihood of future imbalance events (e.g., predict a spike of buy orders incoming). The advantage is a more holistic use of the data, though the complexity is significantly higher.

1.5 Regime-Switching and Structural Models of LOB Dynamics

Another vein of literature approaches high-frequency market dynamics through the lens of *regimes* and structural models that mimic the LOB mechanism. These models acknowledge that market behavior may

change over time, for instance, alternating between calm periods and frenzy periods and that the LOB is essentially a queuing system with strategic agents. Regime-switching models explicitly allow model parameters to change according to a latent state (regime). [16] develops a multi-factor regime-switching model for inter-trade durations in a limit order market. In their framework, the arrival rate of trades (and possibly other order types) switches between different states, such as a high-intensity state and a low-intensity state, driven by underlying factors. They find that such a model fits high-frequency data better than a single-regime (homogeneous) process, capturing the intuition that markets can abruptly shift into faster or slower gears [16]. These regime shifts might correspond to events like news releases or simply endogenous bursts of activity. Importantly for signal extraction, a regime-switching perspective implies that the reliability of signals like OBI could vary by regime, e.g., in a volatile regime, imbalance might need to be interpreted differently than in a steady regime. While Li et al. focus on trade arrival times, the approach can be extended to volumes and imbalances, allowing a richer understanding of how imbalance dynamics behave under different market conditions.

In parallel, researchers have constructed structural models that simulate the mechanics of order books. One influential model is the queue-reactive model proposed by [17]. This model treats the LOB as a collection of queues (one for each price level on each side) and assumes that order arrival and cancellation rates at a given price level depend on the current state of the LOB (such as the queue lengths at that level and neighboring levels). In other words, the placement/cancellation of limit orders is reactive to the existing queue sizes, for example, traders might be more inclined to join a shorter queue (closer to the front of the line for execution) or cancel if their order is too far back in a long queue. By calibrating such a model to data, one can generate synthetic order flow that reproduces realistic patterns (like the distribution of order lifetimes, spread dynamics, etc.). The queue-reactive model provides insights into structural dynamics, such as how imbalances build up and dissipate as a result of many agents' decisions. Recent work by [18] extends this approach by incorporating the *sizes* of orders into the model. In the original formulation, an order was essentially a unit in the queue; Bodor and Carlier recognize that real orders come in varying sizes, which can influence how queues evolve (e.g., a large order at the best bid might deter other buy orders behind it or invite stronger sell competition). Their extended model shows that accounting for order sizes can materially affect the predicted LOB dynamics, thus refining our understanding of how volume contributes to imbalance and liquidity fluctuations [18].

Agent-based models (ABMs) offer another structural perspective by explicitly simulating different types of market participants and their strategies. Instead of a statistical law for event arrivals, an ABM specifies behavior rules for agents (e.g., market makers, momentum traders, HFTs) and lets them interact in a simulated exchange. The study by [8], mentioned earlier for its use of an OBI-based HFT agent, is a prime example of using an ABM to examine structural implications of a strategy. In their model, alongside the OBI-sensitive HFT, there are other agent types (possibly fundamental investors or market makers), and the exchange matches orders in a continuous double auction as in real markets. This approach can recreate complex phenomena like flash crashes or liquidity droughts as emergent outcomes of interactions. Yagi et al. find that when the market is in a stable regime, the presence of an OBI-driven HFT tends to stabilize prices by actively providing liquidity in the direction of imbalance (buying when others are selling, and vice versa), which can dampen short-term volatility [8]. However, during a simulated flash crash scenario, the same OBI strategy did not prevent the crash, importantly, it also did not exacerbate it. This suggests that while OBI-informed HFTs might normally support price discovery and liquidity, their influence can vanish in extreme events when all agents pull back. Such findings contribute to the long-standing debate on whether HFT is a source of market instability or not; the answer may be state-dependent, echoing the idea of regimes.

The evolution of the literature from simple indicators to complex models reflects a growing appreciation of the layered structure in high-frequency markets. Early studies established basic signals like order imbalance and demonstrated their relevance for price prediction. As markets grew faster and more complex, issues of data quality (noise vs signal) prompted efforts to filter and refine these indicators. Concurrently, more sophisticated models, from point processes to agent-based simulations, have been developed to capture the generative processes behind the data. These models offer deeper insights: for instance, a point process model might reveal how an imbalance signal decays over time (via self-excitation decay), whereas a queuing model can show how an imbalance at the top of the book might be quickly

offset by strategic order placement. Each approach, whether empirical or theoretical, contributes to a holistic understanding of filtered versus unfiltered OBI and directional signal quality. Taken together, the surveyed literature highlights that effective signal extraction in high-frequency trading requires both robust data handling (to filter out noise) and robust modeling (to interpret the underlying dynamics). The most recent advances, such as deep learning-based signal construction and hybrid models, continue to push this frontier, aiming to harness the rich information in LOB data while overcoming the challenges posed by its complexity and noise.

Although theoretical frameworks have evolved to capture endogenous dynamics through point process models [11, 12, 9] and regime-switching behavior [16], empirical studies rarely account for heterogeneity in order persistence. Most treatments implicitly assume that all observable events are equally informative. This motivates a fundamental question: *Can directional signals be improved by filtering the LOB based on real-time observable properties such as order survival and update frequency?*

This paper makes four primary contributions to answer that question:

1. **Contemporaneous Correlation Analysis.** We show that filtration enhances the Pearson correlation between raw OBI signals and traded returns. This establishes filtration as an effective tool for improving directional signal clarity at the value level.

2. **Explanatory Power under Discretized Regimes.** We discretize both OBI and returns into categorical regimes and evaluate their relationship using (i) the correlation between regime counts of OBI and the direction of price movement, and (ii) a multivariate regression where return regimes are regressed on OBI regimes. Filtration enhances explanatory power, as quantified by an autoregression-adjusted R^2.

3. **Causal Coherence.** Using a multivariate Hawkes process, we estimate cross-excitation between OBI and return regimes and show that filtration amplifies the directional excitation norm while suppressing spurious self-excitation. This reveals a cleaner causal structure between filtered imbalance and price movements.

4. **Comparative Filter Evaluation.** We implement and benchmark multiple filtration schemes, including lifetime thresholds, modification counts, and modification-time filters, and find that different filters improve different aspects of signal quality across the diagnostic ladder.

We implement a structured empirical pipeline on high-frequency LOB data from BANKNIFTY index futures. The filtering approach targets structurally transient activity by removing orders that reflect fleeting, deceptive, or noncommittal intent. Three distinct mechanisms are evaluated, each grounded in real-time observable features:

1. **Lifetime based filtering**, which discards orders that survive for less than a threshold duration,

2. **Modification count filtering**, which excludes orders subject to frequent updates, and

3. **Modification time filtering**, which retains only those orders whose time between successive updates exceeds a minimum threshold.

Each mechanism refines the event stream to emphasize execution relevant liquidity, and the OBI signal is recomputed on the filtered data accordingly. The directional informativeness of the resulting signal is then evaluated using a layered diagnostic framework.

In addition to filtered and unfiltered imbalance measures derived from order flow updates, we also compute an alternative form of imbalance based exclusively on executed trades. This benchmark, referred to as *OBI from trade events*, is defined as the signed volume difference between market buy and sell executions within each evaluation window. Unlike traditional OBI measures constructed from limit order placements and cancellations, trade-based OBI reflects realized demand and supply as expressed through actual transactions. While this measure is not derived from the visible state of the limit order book, it provides a valuable counterpoint for causal evaluation. By construction, trade-based OBI captures directional conviction that results in execution, and is thus less susceptible to noise introduced by flickering quotes or non-executed strategic activity. As such, it serves as a meaningful reference when assessing the causal coherence of filtered imbalance signals with price movement [19].

The first diagnostic layer computes the contemporaneous Pearson correlation between raw OBI values and traded returns over fixed evaluation windows. This scalar summary offers a baseline measure of

linear association, though it remains symmetric, univariate, and sensitive to noise in high-frequency settings.

The second layer introduces a regime-based representation. Both OBI and returns are discretized into categorical bins and treated as multivariate counting processes. The association is quantified via two complementary diagnostics: the correlation between regime counts of OBI and the direction of price movement, and the autoregression adjusted R^2 from a multivariate regression where return regimes are predicted from OBI regimes. This representation stabilizes fluctuations, compresses extreme values, and transitions the signal toward an event based model structure.

The third layer evaluates causal influence using a multivariate Hawkes process. Regime labeled OBI and return events are treated as a joint point process, and the learned excitation kernel captures the directional influence of past imbalance on future price movements. The submatrix corresponding to OBI to return excitation is extracted, a diagonality weighted mask is applied, and a norm based score is computed to reflect both excitation strength and asymmetry.

Together, the three diagnostic layers isolate complementary dimensions of signal quality. Empirical results demonstrate consistent improvements across all layers under filtered event streams, with the most pronounced gains occurring in the Hawkes excitation scores, highlighting the alignment benefits introduced by structural filtration.

The remainder of this paper is organized as follows. Section 2 introduces the formal problem setting and outlines the structural filtration framework. We define three real-time observable filtering criteria, specify the temporal structure of evaluation and forecast windows, and present a consistent notation for describing the directional scoring objectives.

Section 3 describes the design of the scoring ladder, including the signal construction pipeline, the regime labeling scheme, and the precise definitions of the correlation, regression, and excitation-based diagnostics. Sections 4 and 5 present the empirical evaluation and results. We summarize statistical trends across filters, compare diagnostic scores across models, and demonstrate consistent gains in directional alignment from structurally filtered event streams.

2 Problem Description

We begin by introducing the counting processes and derived statistics used to quantify directional order flow within fixed-length evaluation windows. Each quantity is formally defined and grounded in the structure of raw event sequences extracted from the limit order book.

Formal Definition: Cumulative Event Count.

Let N_t denote the cumulative count of all LOB events observed up to time t:

$$N_t := \sum_{t_i} \mathbb{I}\{t_i \leq t\} \tag{1}$$

where each t_i is an event timestamp and $\mathbb{I}\{\cdot\}$ is the indicator function.

We next define the total number of events arriving in a given lookback window. This interval-level statistic forms the base upon which directional and filtered counts are built.

Formal Definition: Event Count over a Lookback Window.

For fixed horizon h, the number of LOB events in the interval $(\tau - h, \tau]$ is given by:

$$\Delta N_{(\tau-h,\tau]} := \sum_{t_i} \mathbb{I}\{t_i \in (\tau - h, \tau]\} \tag{2}$$

Each event in the sequence carries a directional label $Y_i \in \{b, s\}$ identifying the active side. The next definition isolates directional contributions to the event count.

Formal Definition: Directional Event Count.

The count of events on side $Y \in \{b, s\}$ in the interval $(\tau - h, \tau]$ is:

$$\Delta N^Y_{(\tau-h,\tau]} := \sum_{t_i \in (\tau-h,\tau]} \mathbb{I}\{Y_i = Y\} \tag{3}$$

All standard event types are included without filtering or weighting.

Using these directional statistics, we define a normalized imbalance measure that reflects net order flow pressure within each window.

Formal Definition: Order Book Imbalance.

The order book imbalance (OBI) over the window $(\tau - h, \tau]$ is defined as:

$$OBI(\tau, h) := \frac{\Delta N^s_{(\tau-h,\tau]} - \Delta N^b_{(\tau-h,\tau]}}{\Delta N^s_{(\tau-h,\tau]} + \Delta N^b_{(\tau-h,\tau]}} \tag{4}$$

When the denominator is zero, indicating no directional activity, the interval is excluded from analysis.

To evaluate market response over the same interval, we construct a realized return from executed trades. This requires isolating the relevant subset of the event stream and transforming it into price-based endpoints.

Formal Definition: Realized Return from Trade event Filtration.

Let the LOB event type space be defined as:

$$\mathcal{E} := \{\eta, T, M, C\} \tag{5}$$

where η denotes new order submission, T trade, M modification, and C cancellation.
Let $\epsilon_t \in \mathcal{E}$ denote the event stream. Define a filtration that selects trade events within a lookback window:

$$\mathcal{F}^T_{(\tau-h,\tau]} := \{\epsilon_i \in \epsilon_t : t_i \in (\tau - h, \tau], \ \epsilon_i = T\} \tag{6}$$

Define a transformation that extracts the traded prices at the first and last trade timestamps:

$$\mathcal{G}(\mathcal{F}^T_{(\tau-h,\tau]}) := (p^T_{\tau-}, \ p^T_\tau) \tag{7}$$

where $p^T_{\tau-}$ and p^T_τ are the traded prices at the earliest and latest timestamps in $\mathcal{F}^T_{(\tau-h,\tau]}$, respectively. Then the realized return over the interval is defined as:

$$\tilde{r}_{(\tau-h,\tau]} := \frac{p^T_\tau - p^T_{\tau-}}{p^T_{\tau-}} \tag{8}$$

If $\mathcal{F}^T_{(\tau-h,\tau]} = \emptyset$, the return is undefined and the window is excluded from analysis.

This study does not involve forecasting future returns. Instead, it seeks to evaluate whether the statistical association between order book imbalance and contemporaneous price change, both computed over the same backward-looking interval, is strengthened under filtration of the LOB event stream. Specifically, we assess whether filtering improves explanatory power in terms of (i) empirical correlation between $OBI(\tau, h)$ and $\tilde{r}_{(\tau-h,\tau]}$, and (ii) the structure of fitted excitation kernels in a parametric Hawkes process.

To this end, we introduce a class of filtration schemes that operate on the raw event sequence prior to the computation of any statistics. These schemes remove ephemeral or structurally unreliable events based on short lifetimes, rapid modifications, or flickering placement behavior near the top of book. The resulting filtered event streams yield corresponding statistics, denoted $OBI^{(f)}(\tau, h)$ and $\tilde{r}^{(f)}_{(\tau-h,\tau]}$, which are used for comparative evaluation.

Problem Statement: Evaluation of Filtration-Driven Signal Strength.

Let $\epsilon_t \in \mathcal{E}$ be the observed event stream. Let $\mathscr{F}^{(\cdot)}$ denote a filtration scheme applied to ϵ_t, with superscript $(\cdot)$ indexing a filtering criterion.

Let $\tilde{r}_{(\tau-h,\tau]}$ be the realized return computed over each window as defined previously. Define a scoring functional:

$$\mathcal{S} : \mathscr{F}^{(\cdot)} \times \tilde{r} \longmapsto \mathbb{R} \tag{9}$$

which quantifies the strength of association between filtered order book imbalance and realized return. We consider a set of filtration schemes $\{\mathscr{F}^{(1)}, \ldots, \mathscr{F}^{(k)}\}$, each yielding filtered statistics $OBI^{(i)}$ and a common return sequence $\tilde{r}$. Our objective is to determine whether:

$$\mathcal{S}(\mathscr{F}^{(i)}, \tilde{r}) > \mathcal{S}(\mathscr{F}^{(0)}, \tilde{r}) \quad \text{for some } i \in \{1, \ldots, k\} \tag{10}$$

where $\mathscr{F}^{(0)}$ denotes the identity (unfiltered) case. The specific definitions of $\mathcal{S}$ is provided in subsequent sections.
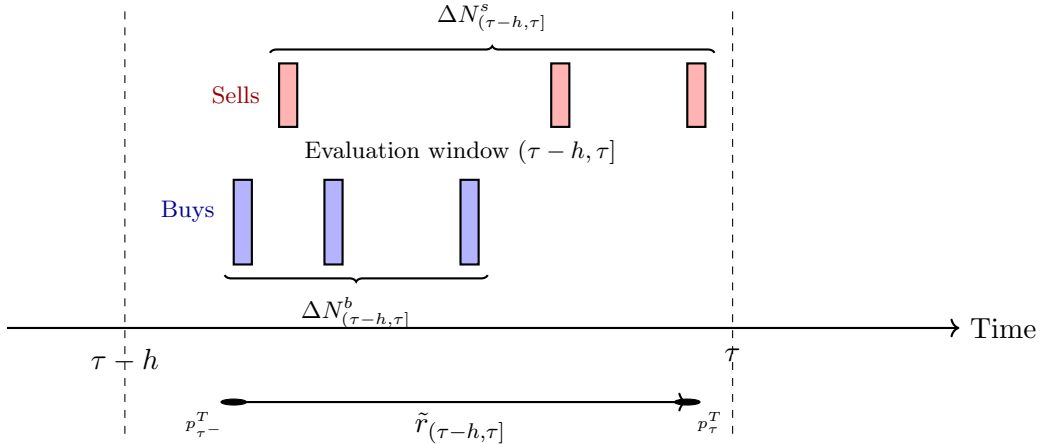

Figure 1: Visual depiction of the evaluation window $(\tau - h, \tau]$ used for computing both order book imbalance and realized return. The imbalance is based on the net directional event counts, while the return is computed using the first and last trade prices within the same window.

3 Methodology

This section formalizes the methodological approach used to evaluate whether structural filtration of limit order book (LOB) event streams improves the directional association between order flow signals and short-term price returns. Our objective is two-fold: first, to assess and calibrate diagnostic scoring methods that quantify directional alignment between Order Book Imbalance (OBI) and realized returns, where OBI serves as a proxy for directional microstructural pressure; and second, to systematically evaluate multiple real-time filtration schemes that aim to isolate execution-viable liquidity by removing short-lived, flickering, or rapidly modified orders from the observed tick stream.

Rather than attempting to forecast future price movements, our focus is on determining whether

9

filtration refines the information content of order flow signals in a way that strengthens their statistical association with contemporaneous traded returns. To this end, we introduce a family of scoring functionals that operate at varying levels of structural abstraction, from raw value-based correlation to regime-based count modeling and causal excitation metrics. These functionals allow us to isolate and quantify improvements in directional clarity arising specifically from filtration, without reliance on model-specific tuning or prediction objectives.

To implement this, we define a family of filtration schemes $\mathscr{F}^{(\cdot)}$ that remove structurally unreliable events from the raw tick stream ϵ_t. Using these filtered events, we reconstruct the LOB and compute directional features over evaluation windows. A reference book constructed from the unfiltered stream provides a baseline. To assess filtration efficacy, we define scoring functionals $\mathcal{S}(\mathscr{F}, \tilde{r})$ that quantify the association between directional features and contemporaneous returns. Additionally, we use the Order Book Imbalance (OBI) indicator to segment and interpret changes in scoring behavior across regimes, though it is not used as an input to the score itself.

3.1   Scoring Framework and Evaluation Philosophy

The central objective of this study is to assess whether structural filtration of the limit order book, by removing short-lived, flickering, or rapidly modified orders, enhances the directional informativeness of flow-based indicators such as Order Book Imbalance (OBI). To this end, we develop a layered scoring framework that evaluates the association between OBI and short-term price returns under progressively richer structural representations. Each layer reflects a different perspective on signal clarity, ranging from raw co-movement to regime-aware causal influence.

The first layer in this framework examines the contemporaneous Pearson correlation between the raw value of the OBI signal and the mid-price return, measured over fixed evaluation windows. This metric is simple, interpretable, and widely used in practice. However, it is also limited: the correlation coefficient is symmetric, univariate, and highly sensitive to transient noise, especially in high-frequency settings where both OBI and returns exhibit significant short-term volatility.

Our primary interest lies in assessing whether the direction and magnitude of OBI are meaningfully associated with the direction and magnitude of returns, specifically, whether higher imbalance values systematically correspond to positive or negative price moves. While contemporaneous correlation offers a starting point, its reliance on pointwise matching between raw OBI and return values neglects structural variation across regimes. Moreover, by operating on raw values without classification, it fails to isolate whether specific levels of pressure, such as extreme buy or sell imbalance, behave differently from neutral states.

A further limitation arises from the way correlation is computed over fixed windows: it implicitly assumes that all observations within a window contribute equally to directional association. This averaging masks the heterogeneity of the underlying flow. In practice, a window may contain a mix of informative and uninformative events, but the correlation measure flattens their contributions. As a result, it may underestimate signal quality in windows where meaningful associations are diluted by high-frequency noise or ephemeral orders.

To overcome these limitations, the second layer shifts from value-based time series to regime-based count representations. Both OBI and return values are discretized into categorical regimes, for example, five bins for OBI and three for returns. Within each evaluation window, we construct count vectors that represent how frequently each regime appears. This transition to count data offers two advantages. It reduces sensitivity to short-lived fluctuations by emphasizing structural composition rather than raw magnitudes, and it aligns the representation more closely with point process models, where event counts naturally define the modeling object. These count vectors allow us to compute a correlation matrix between the occurrence frequencies of OBI and return regimes, revealing how directional pressure aligns with specific return outcomes. Additionally, we perform a multivariate linear regression where the return regime vector is regressed on the OBI regime counts, yielding a coefficient of determination R^2 that reflects the strength of structured association. To filter out artifacts introduced by serial dependence, we fit autoregressive models to each regime sequence, extract the residuals, and recompute the regression. This residualized R^2 isolates meaningful explanatory power beyond autocorrelation. Finally, by repeating this process across multiple lag offsets, we probe the short-term predictive content

of filtered OBI regimes in explaining future return realizations.

The final layer in the scoring framework introduces a causal modeling perspective using a multivariate Hawkes process. Here, the regime-labeled OBI and return events are treated as components of a joint point process. We fit a parametric sum of exponentials excitation kernel, allowing for flexible estimation of how past events influence future occurrences across regime pairs. From the resulting kernel matrix, we extract the submatrix corresponding to OBI-to-return excitation, apply a diagonality-weighted mask to emphasize aligned regime transitions, and compute a norm-based score that captures both the magnitude and structural focus of causal flow. This diagnostic is uniquely suited to capturing the asymmetric, lagged dynamics that structural filtration is expected to clarify.

Raw-value correlation, regime-based alignment, and regime-to-regime excitation form a coherent diagnostic hierarchy. Each layer probes a different aspect of directional informativeness in filtered order flow. The first captures contemporaneous co-movement between imbalance and return. The second assesses whether structured regime transitions align in a statistically meaningful way. The third tests for lagged excitation from imbalance regimes to return regimes, reflecting causal structure. Our empirical results indicate that filtration improves score values across all layers. Correlation-based improvements are the most consistent, followed by moderate gains in regime alignment. Hawkes excitation scores improve only marginally. This pattern suggests that filtration sharpens structural association more reliably than it reveals causal directionality.

3.2 Motivation for Multi-Indicator Scoring.

While structural filtration is designed to isolate meaningful order flow, no single scoring method can fully capture the multiple dimensions of directional informativeness. Raw-value correlation, regime-aligned association, and excitation-based scoring each probe distinct statistical and structural properties of the filtered signal. Correlation captures linear co-movement, but is insensitive to asymmetries and lacks temporal depth. Regression on regime counts introduces structural alignment and interpretability, while Hawkes-based excitation scores reveal causal influence patterns that emerge only once both regime and timing are accounted for. A multi-indicator scoring approach is therefore essential, not for redundancy, but to ensure that improvements in directional clarity are visible across independent, complementary axes of evaluation. We begin by introducing a set of core temporal and structural quantities defined at the order level. Let $t_j^{(1)}$ denote the entry time of order j, defined as the timestamp at which the order is first submitted to the book. Let $t_j^{(2)}$ denote its exit time, defined as the time at which it is either cancelled, fully executed, or replaced. The total time-in-book for order j is given by:

$$\mathcal{T}_j := t_j^{(2)} - t_j^{(1)} \tag{11}$$

This quantity is used to identify ephemeral orders that do not meaningfully interact with the market.

In addition to survival time, we define two modification-based measures. Let M_j denote the number of modifications recorded for order j during its lifetime. We also define $\mathcal{M}_j$ as the time between the last two modifications of order j, provided it was modified at least twice. The quantity $\mathcal{M}_j$ captures timing intensity in the order update stream and serves to flag aggressively re-priced orders.

These quantities form the basis of the filtration schemes presented next.

3.3 Filtration Schemes

The raw event stream observed in high-frequency limit order books reflects a heterogeneous mixture of trading intents, ranging from genuine liquidity provision to strategic misdirection. Orders may be submitted and withdrawn within milliseconds, modified repeatedly, or canceled just before execution, often not to transact but to influence others' behavior. To isolate signals grounded in execution-viable intent, we introduce a family of filtration schemes that remove structurally unreliable events based on real-time observable properties. These filters are designed to suppress the influence of ephemeral and strategically manipulated orders, thereby enhancing the fidelity of derived indicators such as Order Book Imbalance (OBI). Each filtration scheme focuses on a distinct structural marker, order lifetime, modification count, or modification timing, to achieve targeted suppression of noise.

Let $\epsilon_t \in \mathcal{E}$ denote the observed LOB event stream. A filtration scheme $\mathscr{F}^{(\cdot)}$ defines a rule-based removal of structurally unreliable events. We define three filters based on lifetime, modification count, and modification timing.

Formal Definition: Lifetime-Based Filtration.

Let $t_j^{(1)}$ and $t_j^{(2)}$ denote the entry and exit times of order j, and define $\mathcal{T}_j := t_j^{(2)} - t_j^{(1)}$ as its time-in-book. Then:

$$\mathscr{F}^{\mathcal{T}}(\epsilon_t) := \left\{ \epsilon_i \in \epsilon_t : \mathcal{T}_j \geq \bar{\mathcal{T}} \right\} \tag{12}$$

This removes all events corresponding to orders with lifetime below threshold $\bar{\mathcal{T}}$.

While lifetime-based filtering removes orders that vanish too quickly to exert real pressure, it does not address the behavior of persistently modified orders that remain in the book while constantly changing attributes. Orders with high modification counts may reflect strategic probing rather than genuine intent to trade. To suppress such adversarial patterns, we apply a second filter based on the number of post-submission modifications.

Formal Definition: Modification Count-Based Filtration.

Let M_j denote the number of modifications associated with order ID j. Then:

$$\mathscr{F}^{M}(\epsilon_t) := \left\{ \epsilon_i \in \epsilon_t : M_j \leq \bar{M} \right\} \tag{13}$$

This removes events tied to orders with modification count exceeding threshold $\bar{M}$.

While modification count captures the intensity of changes over an order's lifespan, it does not reflect how concentrated those changes are in time. Orders that undergo rapid, successive modifications just before cancellation often serve to spoof or mislead. The final filtration scheme targets such temporally clustered behavior by requiring a minimum separation between the last two modifications, thereby removing orders exhibiting tightly packed adjustment bursts.

Formal Definition: Modification Time-Based Filtration.

Let $\mathcal{M}_j$ be the time between the last two modifications of order j. Then:

$$\mathscr{F}^{\mathcal{M}}(\epsilon_t) := \left\{ \epsilon_i \in \epsilon_t : \mathcal{M}_j \geq \bar{\mathcal{M}} \right\} \tag{14}$$

This removes all events corresponding to orders with tightly clustered final modifications, i.e., $\mathcal{M}_j < \bar{\mathcal{M}}$.

Summary of Filtration Schemes. Each of the three filtration schemes introduced above is applied independently in our analysis, allowing us to isolate the specific structural noise each one targets. The lifetime-based filter removes short-lived orders that are unlikely to represent genuine trading interest. The modification count filter targets orders with frequent quote updates, often indicative of speculative or non-committal behavior. The modification time filter excludes orders with rapid-fire adjustments near the end of their lifetime, a pattern frequently linked to spoofing tactics. By treating these filters separately, we assess their individual contributions to signal stability and directional clarity without confounding their effects.

3.4   Evaluation Indicators: Order Book Imbalance (OBI and OBI$^{(\text{T})}$)

To support interpretation of score behavior across filtration schemes, we employ the Order Book Imbalance (OBI) as an auxiliary, regime-defining indicator. Although not directly used in scoring, OBI provides a granular, tick-level measure of passive liquidity pressure that shapes directional signal clarity under various filters.

OBI is a well-established measure of passive liquidity imbalance in market microstructure literature. Empirical studies demonstrate that OBI is strongly associated with short-term price pressure and alpha generation [20, 21, 4]. Its tick-level resolution allows seamless integration with Hawkes-based excitation analysis, while its ability to be averaged over fixed windows supports contemporaneous correlation frameworks.

Order Book Imbalance. We use the formally defined Order Book Imbalance (OBI) measure introduced in Equation (4), which captures the directional skew in event arrivals over a backward-looking window $(\tau - h, \tau]$. This definition is based on the net difference between arrival counts of sell-side and buy-side events, normalized by their total. It is particularly suited to our filtered event streams, where transient activity can be excluded before imbalance computation.

OBI values close to ± 1 indicate extreme imbalance and are often predictive of price movement direction. We compute OBI from the book reconstructed after each filtration scheme and use it to segment return-score behavior into distinct imbalance regimes, highlighting where structural filtering yields the greatest directional clarity.

While standard OBI provides a real-time view of *latent* book pressure, it is susceptible to noise from fleeting orders and unexecuted strategic placements. To construct a cleaner, execution-grounded benchmark, we introduce an alternative imbalance measure derived solely from signed trade activity. This trade-based OBI captures *realized* directional pressure by tallying buyer-initiated versus seller-initiated transactions within each window. The resulting signal is immune to book flicker and better reflects market commitment. We now define this benchmark formally.

Formal Definition: Order Book Imbalance by Trades ($\mathbf{OBI^{(T)}}$).

Let $N_{[\tau-h,\tau)}^{(b,T)}$ and $N_{[\tau-h,\tau)}^{(s,T)}$ denote the total number of trades initiated by buyers and sellers, respectively, within window $[\tau - h, \tau)$. Then:

$$\mathrm{OBI}^{(T)}(\tau) := \frac{\Delta N_{[\tau-h,\tau)}^{(b,T)} - \Delta N_{[\tau-h,\tau)}^{(s,T)}}{\Delta N_{[\tau-h,\tau)}^{(b,T)} + \Delta N_{[\tau-h,\tau)}^{(s,T)}}$$

Trade-based OBI differs fundamentally from standard book-derived imbalance. While traditional OBI is computed from standing volume at the best bid and ask levels, capturing *latent* liquidity. The trade-based version relies exclusively on *realized* executions. Specifically, we follow the methodology proposed in [19], where each trade is signed using a tick-based classification rule, and imbalance is computed as the net signed trade count within each evaluation window. This construction eliminates noise from flickering quotes, non-executed strategic placements, and ephemeral cancellations. Although it is not a filtered signal in the same sense as our structural filtration schemes, it provides a clean and grounded benchmark for directional conviction. As such, $\mathrm{OBI}^{(T)}$ is used in our analysis as a reference point when evaluating causal alignment via Hawkes excitation norms.

3.5 Scoring Functionals

To evaluate the directional informativeness of filtered order flow, we introduce a sequence of scoring functionals, each designed to quantify the strength and structure of association between directional signals, such as order book imbalance (OBI), and contemporaneous returns. The functionals are ordered by increasing structural sophistication, beginning with simple empirical association and culminating in event-level causal excitation analysis.

We begin with *Pearson correlation*, a familiar and computationally lightweight metric. It serves as a high-level diagnostic of whether filtered directional features track return fluctuations more closely than their unfiltered counterparts. However, correlation flattens out nuances by requiring time-synchronous aggregation and discards structural details that may lie in return regimes or flow asymmetries.

To address this, we next employ a *discretized regime structure score*. Here, both the returns and the directional signals are discretized into categorical regimes. By computing the coefficient of determination

R^2 between indicator regime frequencies and return regimes within each sampling window, we capture whether specific OBI states are systematically linked to directional price moves. This formulation eliminates the need for fixed lags and emphasizes localized structural associations.

Building on this, we introduce a *Hawkes-based scoring functional* that models the return-OBI interaction as a multivariate point process. Unlike correlation or regression, this method accounts for *asymmetric causal excitation* between return regimes and flow regimes. The excitation kernels ϕ_{ij} describe how the occurrence of one type of event (e.g., an extreme OBI state) increases the likelihood of another (e.g., a sharp return). The learned excitation matrix is post-processed using a diagonally-weighted mask to quantify the clarity of directional causal flow, a novel contribution in this context.

Taken together, these scoring functionals form a layered evaluation framework. Each level allows the analyst to ask progressively finer questions. The comparison between correlation and discretized regime structure reveals whether a small subset of return regimes dominates the association signal. The jump from regression to Hawkes excitation exposes whether such associations are symmetric or directional, for example, whether extreme returns follow imbalance, or whether they excite further flow. This diagnostic ladder ultimately supports a more grounded interpretation of signal quality under each filtration scheme.

Each functional quantifies the association between returns and structural features derived from the order flow. These scores do not use the Order Book Imbalance (OBI) directly as an input but are later interpreted by conditioning on OBI regimes to assess the signal stability under varying liquidity pressure.

(1) Contemporaneous Correlation Score. The first diagnostic layer assesses the contemporaneous Pearson correlation between the raw value of the Order Book Imbalance (OBI) signal and traded returns, evaluated over fixed-length time windows. This correlation coefficient provides a coarse, symmetric measure of linear association. It does not capture temporal directionality, nor does it distinguish between causal and coincidental alignment. Its primary utility lies in offering an interpretable scalar summary of alignment between signed pressure and subsequent price movement.

However, in the high-frequency setting, this diagnostic faces structural limitations. Both OBI and returns are highly volatile and sensitive to transient flow. The raw correlation operates on pointwise value matches, assuming equal informational weight across all observations within a window. This masks intra-window heterogeneity, blurring the distinction between structurally persistent pressure and ephemeral bursts. Moreover, by treating the OBI signal as a continuous series of values rather than as a categorical indicator of directional stance, the correlation score fails to reveal whether extreme pressure regimes behave differently from neutral ones.

As such, this layer serves primarily as a baseline. Its purpose is not to predict future returns, but to calibrate the association strength of raw imbalance values with contemporaneous price changes under different filtration schemes. Any observed improvement in this score under filtering provides early evidence that structurally unreliable events were distorting the unfiltered signal.

The first scoring functional, $\mathcal{S}^\rho$, computes the Pearson correlation between OBI and realized traded returns averaged over evaluation windows. This provides a scalar diagnostic of linear association, intuitive for practitioners and informative for real-time signal clarity assessment.

Formal Definition: Contemporaneous Correlation Score.

Let $X = \mathrm{OBI}(\tau)$ and $Y = \tilde{r}_\tau$ denote the averaged OBI and return over window $[\tau - h, \tau)$. Then:

$$\mathcal{S}^\rho(\mathscr{F}, \tau) := \rho(X, Y) = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sqrt{\mathbb{V}[X] \cdot \mathbb{V}[Y]}}$$

where expectations are computed over a rolling ensemble of evaluation windows.

This method provides a high-level summary of the relationship between filtered OBI and returns, but does not capture nonlinear or regime-specific behavior. Moreover, correlation is symmetric and non-directional, making it agnostic to causal flow.

(2) Explanatory Power under Discretized Regimes To overcome the limitations of value-based correlation, the second diagnostic layer transforms both OBI and return series into categorical

regimes and analyzes their structural alignment through count-based representations. Each evaluation window is segmented into regime bins, for example, nine bins for OBI and four for returns, and count vectors are constructed to capture the frequency with which each regime occurs.
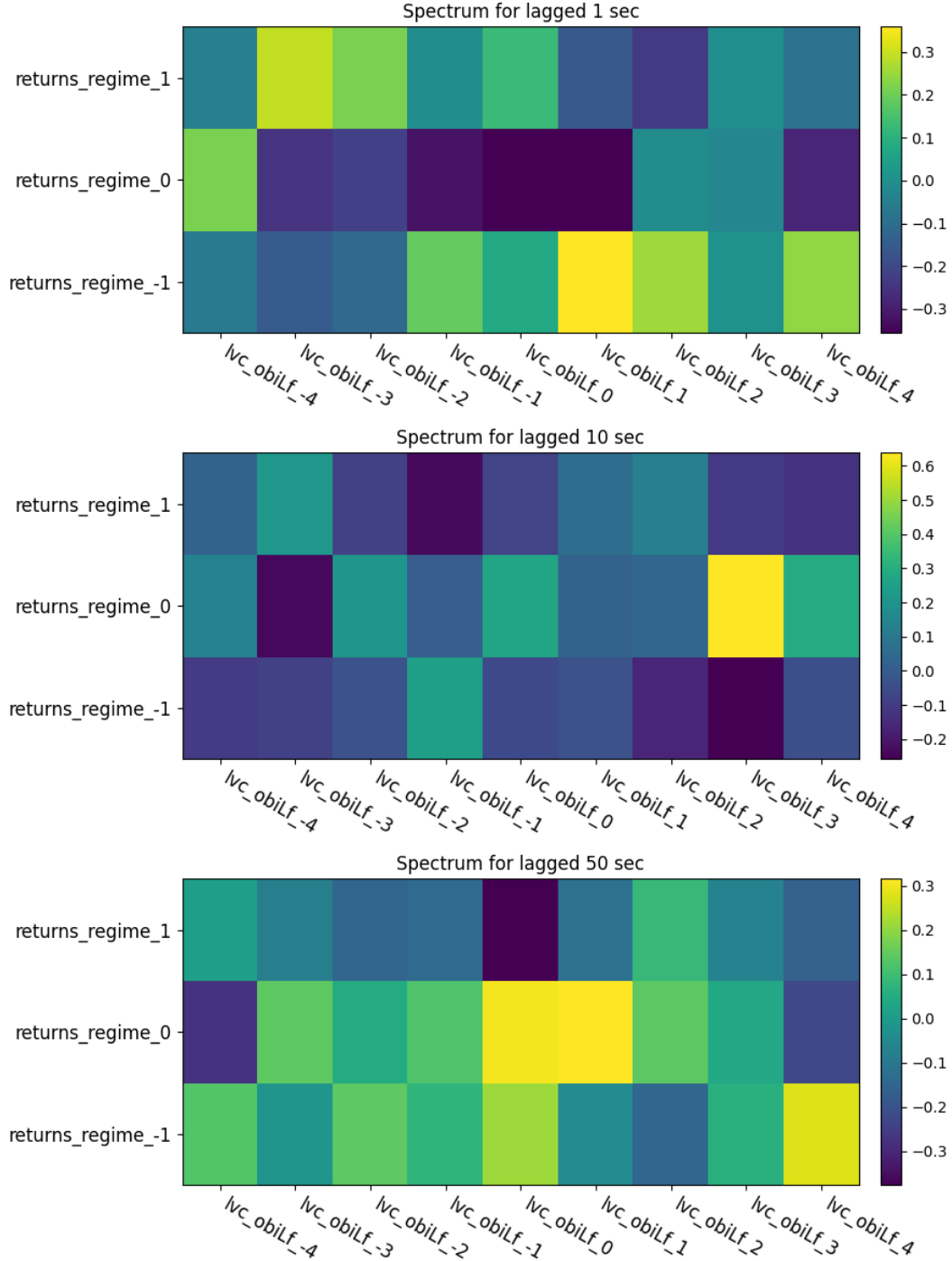

Figure 2: Discretization of OBI and returns into regime bins. Each evaluation window is mapped to a count vector over OBI regimes and a one-hot encoded return regime, forming the basis for correlation and regression analysis under a point-process framework.

This transformation introduces two key benefits. First, it reduces noise sensitivity by abstracting away from raw magnitudes and instead emphasizing the structural composition of directional pressure. Second, it moves closer to a point process representation, where regime counts are natural units for modeling and regression.

Within this discretized regime framework, we evaluate explanatory power through two association mechanisms. The first involves computing a matrix of Pearson correlation coefficients between OBI regime counts and return regime indicators. The second fits a multivariate linear regression, regressing return regime counts on OBI regime counts to obtain a window-specific coefficient of determination R^2.

To control for spurious correlation due to serial dependence, we fit autoregressive models to the regime sequences and recompute the regression on the residuals. This residualized R^2 isolates genuine structural association. Finally, by repeating the analysis over several lags, we evaluate whether filtered OBI regime counts carry explanatory power for future return regimes, thereby quantifying the directional persistence introduced by filtration.

Let the OBI process be discretized into 9 regimes and the return process into 4 regimes, corresponding to quantized bins within $[-1, +1]$ based on observed support.

For each window $[\tau - h, \tau)$, we construct two vectors: one that counts how many times each OBI regime appears (across nine bins), and another that records which of the four return regimes is active by using a one-hot encoded indicator.

To assess directional alignment in this regime space, we define two complementary scoring functionals. The first is a diagonality-weighted correlation score, $\mathcal{S}^{\rho, \Lambda}$, which captures the average Pearson correlation between OBI and return regime counts across all evaluation windows. For each window τ, we compute a regime-regime correlation matrix $\rho_\tau \in \mathbb{R}^{9 \times 4}$, and aggregate these matrices using a diagonality-emphasizing mask $\Lambda \in \mathbb{R}^{9 \times 4}$. The score is defined as the masked ℓ_1-norm of the average correlation matrix:

$$\mathcal{S}^{\rho, \Lambda}(\mathscr{F}) := \left\| \Lambda \circ \left(\frac{1}{B} \sum_{\tau \in \mathcal{T}} \rho_\tau \right) \right\|_1 \tag{15}$$

The second functional, $\mathcal{S}^{\mathcal{R}}$, quantifies how well OBI regime counts explain return regime structure via multivariate linear regression. We regress $\mathbf{R}_\tau$ on $\mathbf{Q}_\tau$ via ordinary least squares:

$$\mathbf{R}_\tau = \beta \mathbf{Q}_\tau + \varepsilon_\tau$$

The scoring functional $\mathcal{S}^{\mathcal{R}}$ is then defined as the coefficient of determination R^2 of this regression, aggregated over windows:

$$\mathcal{S}^{\mathcal{R}}(\mathscr{F}) := \sum_{\tau \in \mathcal{T}} R^2(\tau) \tag{16}$$

> **Modeling Remark.**
> Regime binning is performed symmetrically around zero, with uniformly spaced intervals across $[-1, +1]$, and a central neutral regime for both OBI and returns. The masked correlation score $\mathcal{S}^{\rho, \Lambda}$ emphasizes diagonality, favoring aligned pairings of directional pressure and return outcome, while discounting diffuse or off-regime associations. To evaluate short-term predictive alignment, we repeat the regression-based score $\mathcal{S}^{\mathcal{R}}$ across multiple lags ℓ, where OBI at time τ explains returns at $\tau + \ell$. To remove spurious effects from autocorrelation, we fit autoregressive (AR) models to both series and recompute scores using the residuals. Results are reported both with and without this correction to highlight the robustness of directional association under filtration.

(3) Hawkes Excitation Norms   The final layer in the scoring framework adopts a causal modeling perspective using a multivariate Hawkes process. Regime-labeled OBI and return events are treated as components of a joint point process, and we estimate their excitation interactions with a parametric sum of exponentials kernel form. This yields a full excitation kernel matrix that encodes how the occurrence of events in each OBI regime influences the arrival intensity of return events across all return regimes.

To isolate the directional component of interest, we extract the submatrix that maps OBI regimes to return regimes, thereby focusing on the forward causal influence of filtered order flow. To further emphasize alignment between structurally similar regimes, we apply a diagonality-weighted mask that amplifies contributions from aligned OBI-return regime pairs while suppressing off-diagonal excitation. The Frobenius norm of this masked submatrix is then computed as a scalar diagnostic score. This final metric captures both the magnitude and directional precision of the excitation structure and is especially well suited to detecting latent causal interactions clarified by structural filtration.

Specifically, from the fitted kernel matrix $\Phi \in \mathbb{R}^{13 \times 13}$, we extract the submatrix $\Phi_{\mathrm{OBI} \to \mathrm{Ret}} \in \mathbb{R}^{4 \times 9}$, isolating causal influence from OBI regimes to return regimes. Using a diagonally-weighted mask M to emphasize regime-matched excitation strengths, the scoring functional is defined as:

$$\mathcal{S}^{\phi}(\mathscr{F}) := \|\Phi_{\mathrm{OBI} \to \mathrm{Ret}} \circ M\|_1 .$$

This score measures directional causality, whether specific imbalance regimes tend to trigger subsequent return regimes. By excluding within-group excitations (OBI–OBI or return–return), we focus on cross-regime signal transfer. The combination of Hawkes modeling and sum of exponentials kernels provides a deep, structure-rich lens for assessing the impact of filtration on directional flow dynamics.

> **Modeling Remark.**
> Only excitation terms representing cross-regime interactions between OBI and return regimes are retained. Self-excitation among OBI regimes and among return regimes is excluded. This score isolates true directional signal transfer. Additionally, all scoring functionals are computed over identical evaluation windows across filtration schemes to ensure any observed improvements arise purely from the filtration itself and not from sampling or aggregation artifacts.

(4) OBI by trades In addition to the imbalance signals derived from LOB state under various filtration schemes, we also evaluate a structurally distinct signal constructed purely from executed trades. This benchmark, referred to as *OBI from trade events*, measures the signed volume difference between buyer-initiated and seller-initiated trades within each evaluation window. Unlike standard OBI, which is computed from standing book volume or order flow updates, this variant reflects realized market pressure through transaction volume. It is not subject to flickering quotes or ephemeral cancellations, and provides a reference point for assessing causal alignment. Prior work has shown that trade-based OBI exhibits stronger causal coherence with future price movements [19]. While it is not a filtered signal per se, it serves as a valuable benchmark for interpreting the effectiveness of structural filtration in revealing causal excitation patterns.

Summary of Scoring Functionals. The layered structure of our scoring framework offers a coherent diagnostic ladder for evaluating directional informativeness. The first layer assesses value-level co-movement using raw Pearson correlation. The second captures structural alignment through regime-discretized counts and explanatory regression. The third models causal excitation using point process dynamics. Together, these layers reveal how filtration clarifies directional signals by reducing noise, emphasizing regime-level structure, and uncovering causal influence.

To maintain strict comparability across filtration schemes, all scoring functionals are computed using identical evaluation windows and normalization logic. For each variant, we reconstruct the limit order book from the filtered event stream, extract directional signals using the same conventions, and apply the scoring pipeline without model-specific tuning. This ensures that any observed differences in association strength can be attributed solely to structural filtering, without confounding effects from reparameterization or inconsistent return handling.
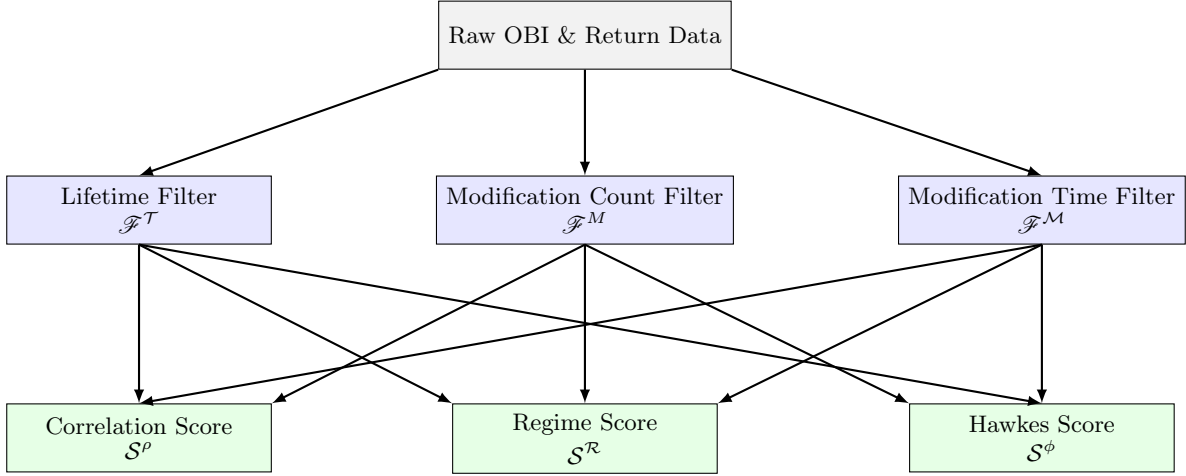
Figure 3: Schematic overview of the filtration and scoring pipeline. Raw order book imbalance and return streams are processed through three distinct filtering schemes. The resulting filtered data are then evaluated using correlation, regime-based, and Hawkes excitation scoring functionals. This layered structure allows a modular assessment of signal quality under different microstructure filters.

4 Data

Our empirical study is based on tick-by-tick data for BANKNIFTY futures and selected equities sourced from the Indian National Stock Exchange (NSE). Each tick records a discrete event involving one of four categories: new order submission, order modification, order cancellation, or trade execution. Every order is uniquely identified by an Order ID (OID), which remains associated with all its subsequent modification events until the order is either cancelled or executed.

Each tick includes a market snapshot comprising the top five levels of the limit order book (LOB), including bid/ask prices and corresponding standing volumes. Importantly, the dataset is *event-driven*: only those ticks that induce a change in the top-5 levels of the book are recorded. As a result, periods of inactivity or activity beyond the top 5 levels are omitted from the stream, yielding a compact, latency-sensitive representation of market activity.

To analyze how filtering affects the directional information content of the order flow, we reconstruct alternate versions of the limit order book under different filtration schemes. The filtering process is implemented as follows:

1. For each filtration scheme $\mathscr{F}^{(\cdot)}$, we construct a list of Order IDs (OIDs) that are to be excluded from contributing to the displayed book.

2. During reconstruction, any new or modified order matching an excluded OID is ignored, except in the case of trade execution. If a filtered OID executes against a resting order, its trade tick is retained to preserve execution consistency and prevent structural imbalance in the reconstructed book.

This protocol ensures that downstream scoring functionals operate over books that remain structurally valid and comparable across filtration schemes.

5 Experimental Setup

This section presents the empirical evaluation of directional signal extraction via order book filtration. Our goal is not to forecast future returns, but to assess whether short-term return association is stronger when order flow signals are filtered using well-defined structural attributes. We consider three scoring functionals: correlation, discretized regime structure, and Hawkes excitation norms. These are applied to two signal types: filtered order book imbalance (OBI) and OBI computed using trade events.

We evaluate our methodology using tick-level BANKNIFTY futures data from the National Stock Exchange of India on three distinct trading dates: 2nd January 2021, 13th January 2021, and 24th January 2021. Each event record includes server timestamp, order ID, event type (NEW, MODIFY, CANCEL, TRADE), side (BID/ASK), price, and quantity. Book reconstruction was performed at millisecond precision to extract limit order book (LOB) states.

Evaluation windows of length $h = 10$ seconds were anchored every 15 seconds over the main trading hours (09:20–15:25), with the forecast window $(\tau, \tau + \xi]$ of length $\xi = 1$ second used solely to compute the realized return $\tilde{r}$.

We construct filtered versions of the event stream ϵ_t using three criteria as defined in Section 3.3:

- $\mathscr{F}^{\mathcal{T}}(\epsilon_t)$: Lifetime-based filter retaining only events with total survival time $\mathcal{T}_j \geq \bar{\mathcal{T}}$.

- $\mathscr{F}^M(\epsilon_t)$: Modification count filter retaining events with $M_j \leq \bar{M}$.

- $\mathscr{F}^{\mathcal{M}}(\epsilon_t)$: Modification-time filter retaining events where first modification lag $\mathcal{M}_j \geq \bar{\mathcal{M}}$.

Each filtered stream is used to compute order book imbalance (OBI) as the directional signal over $[\tau - h, \tau)$, and evaluated against the realized return $\tilde{r}$ over the forecast window. In addition, we compute OBI using trade events alone (henceforth called trade-based OBI), and apply the Hawkes-based causal scoring functional to evaluate its excitation structure.

The evaluation is repeated across different thresholds:

- Lifetime thresholds $\bar{\mathcal{T}} \in \{100, 500, 1000\}$ ms

- Modification lag thresholds $\bar{\mathcal{M}} \in \{50, 100, 200\}$ ms

- Modification count thresholds $\bar{M} \in \{1, 3, 5\}$

While the experimental setup establishes a unified scoring framework and consistent evaluation structure across multiple trading days, the results reveal a nuanced picture. Associative scores, particularly correlation and discretized regime cross-correlation, show tangible improvement under structural filtration, with modification-time filtering offering the most consistent gains. However, improvements in causal excitation, as captured by Hawkes norm scores, are comparatively modest when OBI is derived from limit order book events. In contrast, when OBI is computed using trade events alone, excitation intensity improves markedly. This contrast between associative clarity and causal alignment serves as the foundation for the sections that follow.

6 Results

We present our empirical findings regarding the effectiveness of filtration schemes applied to Order Book Imbalance (OBI) and their impact on directional signal extraction in high-frequency BANKNIFTY futures data. In addition to the standard book-based OBI, we also evaluate an alternative signal derived solely from trade executions, denoted $\mathrm{OBI}^{(\mathrm{T})}$. Three scoring functionals are evaluated: the Correlation Score $\mathcal{S}^\rho(\mathscr{F}, \tau)$, the Discretized Regime Score using Cross-Correlation, and the Discretized Regime Score using R^2. In addition, causal intensity is assessed via the Hawkes Excitation Norm Score $\mathcal{S}^\phi(\mathscr{F})$. These layers jointly assess linear association, regime-level explanatory strength, and causality.

The experiments span three representative trading days, January 2, 13, 23 in 2021, and three real-time filtration schemes: lifetime-based ($\mathscr{F}^{\mathcal{T}}$), modification count ($\mathscr{F}^M$), and modification-time ($\mathscr{F}^{\mathcal{M}}$). Each filtration is benchmarked against unfiltered OBI.

To synthesize the impact of filtration across all evaluation layers, Table 1 presents the average scores for each filtration scheme and scoring functional. These values are computed across three trading days using standard OBI, with autocorrelation removed for the Pearson and discretized regime-based association scores. For causal inference, the Hawkes Excitation Norm Score is estimated using the sum of exponentials kernel variant applied uniformly across filters.

Table 1: Summary of Averaged Scores Across Scoring Functionals and Filtration Schemes

Filter Type	$\mathcal{S}^{\rho}(\mathscr{F})$	$\mathcal{S}^{\rho,\Lambda}(\mathscr{F})$	$\mathcal{S}^{\mathcal{R}}(\mathscr{F})$	$\mathcal{S}^{\phi}(\mathscr{F})$
Unfiltered (UF)	0.01018	-4.00	8.43	8.9292
Lifetime (LF)	0.01011	-3.41	8.39	8.9287
Mod Count (MF)	0.00826	-4.44	8.42	9.0048
Mod-Time (MTF)	0.01133	-3.39	8.37	9.1745

Table 2: *Note*: All values are averaged across the three trading days. Pearson and discretized regime scores are computed after removing auto-correlation. Hawkes score uses the sum of exponentials kernel variant. See Appendix Tables 5, 7, and 8 for full breakdowns.

6.1 Key Observations

Linear Correlation ($\mathcal{S}^{\rho}$): The modification-time filter (MTF) yields the highest improvement in Pearson correlation (0.01133) over the unfiltered score (0.01018), a gain of approximately 11.3%. Lifetime filtering shows near-parity with baseline (0.01011), while modification-count filtering underperforms (0.00826).

Discretized Regime Scoring (Cross-Correlation): Here too, the modification-time filter (MTF) achieves the strongest improvement, reducing the average regime-wise cross-correlation error from -4.00 (UF) to -3.39. Lifetime filtering shows moderate gains (-3.41), while modification-count filtering slightly worsens performance (-4.44).

Discretized Regime Scoring (R^2): Performance is relatively flat across filters. The unfiltered score is 8.43, with filtered variants yielding similar results: 8.39 (LF), 8.42 (MF), and 8.37 (MTF). No consistent pattern of improvement is observed.

Hawkes Excitation Norm Score ($\mathcal{S}^{\phi}$): Causal excitation intensity improves modestly under all filtration schemes. MTF yields the highest score (9.1745 vs. 8.9292 UF), followed by MF (9.0048), while LF (8.9287) shows minimal change.

6.2 Trade-Based OBI Results

In addition to traditional book-derived OBI, we evaluate an alternative imbalance measure $OBI^{(T)}$ based exclusively on signed trade activity. This formulation, defined and justified in Section 3.1, constructs the imbalance signal from realized executions rather than standing book volume. The same filtration schemes are applied, and the resulting Hawkes excitation norms are reported in Appendix Table 9.

Unlike standard OBI, the trade-based formulation reveals a more pronounced variation across filters. For instance, the excitation norm under modification-time filtering (MTF) rises sharply to 24.7352 on January 23, compared to 9.6726 for standard OBI on the same day. Lifetime and modification-count filters also exhibit substantial gains (e.g., 15.8630 on January 13 under LF).

This result indicates that when OBI is constructed from trade-induced order flow, the signal is far more responsive to structural filtration. Specifically, the directional clarity extracted under MTF and LF translates into significantly higher causal excitation in return processes.

6.3 Summary Interpretation

Filtering OBI based on observable event structure yields tangible improvements in associative strength, particularly under Pearson correlation and regime-wise cross-correlation. The modification-time filter (MTF) is consistently dominant across these layers. However, gains in causal excitation, as measured by Hawkes norms, remain modest when applied to standard OBI.

By contrast, when OBI is reconstructed from trade events alone, filtration yields markedly higher causal excitation scores, particularly under MTF and LF. This suggests that causal pathways from OBI to returns are more active when the imbalance signal is derived from executed trades, rather than inferred from book composition. In high-frequency settings where order-level noise and quote flickering obscure latent intent, realized flow may provide a more robust anchor for causality-based inference.

7   Discussion

Our findings offer a nuanced view of how structural filtering affects the informational quality of high-frequency order flow signals. When OBI is computed from limit order book state alone, filtration enhances associative metrics, such as Pearson correlation and cross-regime alignment, but exhibits limited efficacy in strengthening causal linkages to contemporaneous returns. This divergence between associative and causal scoring layers highlights a key tension in high-frequency alpha design.

The strongest associative improvements are observed under modification-time filtering, which removes ephemeral events that fail to persist even briefly within the book. This filtering enhances signal clarity without requiring knowledge of future order modifications or cancellations, making it realistic for real-time implementation. Across all three trading days, MTF consistently improves correlation scores and discretized regime alignment, underscoring its robustness as a pre-processing step for OBI computation.

However, the same filtering does not yield comparable gains in Hawkes-based excitation norms. Even when applied to carefully curated OBI signals, the causal excitation from OBI to returns remains largely unchanged. This is most evident in the Sum of exponential kernel norms, where the differences between filtered and unfiltered OBI hover within a narrow band. The limited rise in excitation intensity suggests that OBI, as derived from book state, may lack sufficient alignment with the latent triggers of high-frequency return movements.

To understand this further, we examine the structure of the fitted Hawkes kernels. The cross-excitation components, from OBI to returns, are consistently weak across all filters. By contrast, self-excitation within OBI and returns remains dominant. This dominance of self-excitation suggests that both OBI and returns exhibit internal memory, where past activity reinforces future behavior. In the case of OBI, this may reflect microstructural feedback mechanisms such as quote stuffing or passive order replenishment. For returns, strong self-excitation aligns with short-term momentum bursts often observed in high-frequency price dynamics. The relative weakness of cross-excitation, when contrasted against this pronounced self-dependence, implies that directional signals may be driven more by internal clustering than by causal flow-through from one process to another. This asymmetry implies that the observed predictability may stem more from endogenous clustering (e.g., momentum or quote stuffing effects) rather than causal influence. In effect, the order book appears to reflect a self-reinforcing structure, where returns and order flow evolve independently yet auto-correlate within their domains. From a strategy design perspective, this favors momentum-based approaches that exploit short-term clustering, rather than mean-reverting strategies that rely on exogenous flow shocks. The endogenous excitation of returns suggests that price moves tend to reinforce themselves, independent of contemporaneous OBI inputs. In such regimes, causal inference from filtered order flow may yield limited benefit unless paired with terminal event streams such as executed trades.

The contrast becomes sharper when OBI is redefined using trade events alone. In this setup, causal excitation norms increase dramatically under filtration. Notably, under MTF filtering, the excitation norm more than doubles compared to regular OBI on January 23. This suggests that trade-induced OBI captures a stronger causal footprint, likely because trades are terminal events that register true execution pressure rather than intent. This distinction has important implications for excitation modeling. Since standard OBI reflects transient quote activity, it may lack alignment with actual return-generating mechanisms, resulting in weak and noisy cross-excitation paths. In contrast, $OBI^{(T)}$ anchors directional inference on terminal execution events, enabling more precise and recoverable causal structure within the fitted Hawkes kernels. Filtering amplifies this by excluding short-lived or non-informative trades, thereby isolating directional shocks more effectively.

Activity levels across the three days support this interpretation. January 2 exhibits relatively muted

excitation behavior, consistent with low event intensity. On January 23, a high-activity day, trade-based OBI filtered with MTF produces the sharpest jump in causal norm. This heterogeneity illustrates how event-driven filtration interacts with baseline liquidity conditions, with greater returns to filtering arising under high market stress or volatility.

Overall, our results suggest that filtration is highly effective in improving associative signal clarity, especially for correlation-based measures. However, when the objective is to extract causally coherent signals, the nature of the underlying event stream becomes critical. Order book state, even when filtered, may not provide sufficient causal alignment. In contrast, trade-based OBI, when coupled with structural filtration, reveals latent excitation structure more clearly.

Beyond empirical contrast, the performance of trade-based OBI also carries methodological implications. The marked increase in causal excitation norms under filtering suggests that $OBI^{(T)}$ enables more precise inference of directional structure compared to standard book-derived measures. Since trades reflect realized execution rather than intent, the excitation kernels estimated from filtered trade-based OBI exhibit stronger identifiability and sharper cross-excitation structure. This makes $OBI^{(T)}$ not only a superior diagnostic for causal footprints, but also a more reliable input for strategy design in high-frequency environments where predictive coherence and structural recoverability are essential.

8   Conclusion

This study introduces a structured methodology for evaluating directional signal quality under real-time order flow filtration in high-frequency markets. By integrating three layers of diagnostic scoring: linear correlation, discretized regime alignment, and Hawkes-based causal excitation, we provide a unified framework to assess the associative and causal properties of signals derived from limit order book data.

While our empirical analysis is restricted to three trading days of BANKNIFTY futures data, the experimental results highlight several patterns. Most notably, filtering order events based on observable structural properties, such as modification count, lifetime, or update intervals, improves the clarity of Order Book Imbalance (OBI) as a contemporaneous directional signal. The most consistent gains are observed under modification-time filtering, particularly in correlation and discretized regime-based scores. However, these improvements do not extend uniformly to causal excitation scores: across all filter types, the causal linkage from filtered OBI to returns remains weak.

The contrast becomes significant when OBI is instead derived using trade events. Here, causal scores exhibit marked improvements under filtering, suggesting that execution-based measures better reflect directional impact mechanisms. This reinforces a practical insight: while filtered OBI from order book state may capture short-term momentum or clustering effects, trade-based OBI may serve as a more viable proxy for predictive signal generation.

We note that the empirical findings presented here reflect the behavior of filtered signals within the chosen dataset and experimental setup. Rather than asserting universal market microstructure conclusions, our aim is to demonstrate how a multi-layered diagnostic framework can be applied to interrogate the associative and causal structure of directional signals under event filtration. The methodology is extensible to broader settings and may serve as a template for future investigations into the structural drivers of high-frequency alpha performance.

References

[1] Michael Hanke and Markus Weigerding. Order flow imbalance effects on the german stock market. *Business Research*, 8(2):213–238, 2015.

[2] Petter N. Kolm, Jorge Turiel, and Nicholas Westray. Deep order flow imbalance: Extracting alpha at multiple horizons from the limit order book. *Mathematical Finance*, 33(4):1044–1081, 2023.

[3] Rama Cont, Arseniy Kukanov, and Sasha Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88, 2014.

[4] Joel Hasbrouck and Gideon Saar. Low-latency trading. *Journal of Financial Markets*, 16(4):646–679, 2013.

[5] Maximilian Ulze, Thomas Stadler, and Andreas Rathgeber. The case of fleeting orders and flickering quotes. https://doi.org/10.2139/ssrn.3935954, 2021. SSRN Working Paper.

[6] Eric Budish, Peter Cramton, and John Shim. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621, 2015.

[7] Petter Dahlström, Björn Hagströmer, and Lars L. Nordén. The determinants of limit order cancellations. *The Financial Review*, 59(1):181–201, 2024.

[8] Itsuki Yagi, Masaru Hoshino, Toshihiro Mizuta, and Nan Cao. Impact of high-frequency trading with an order book imbalance strategy on agent-based stock markets. *Complexity*, 2023:1–14, 2023.

[9] Clive G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.

[10] Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[11] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(1):1550005, 2015.

[12] Stephen J. Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. Critical reflexivity in financial markets: A hawkes process analysis. *The European Physical Journal B*, 86(10):442, 2013.

[13] Vladimir Filimonov and Didier Sornette. Apparent criticality and calibration issues in the hawkes self-excited point process model: Application to high-frequency financial data. *Quantitative Finance*, 15(8):1293–1314, 2015.

[14] Marcello Rambaldi, Emmanuel Bacry, and Fabrizio Lillo. The role of volume in order book dynamics: A multivariate hawkes process analysis. *Quantitative Finance*, 17(7):999–1020, 2017.

[15] Sobin Joseph and Shashi Jain. A neural network based model for multi-dimensional nonlinear hawkes processes, 2023. Preprint.

[16] Zhicheng Li, Haipeng Xing, and Xinyun Chen. A multifactor regime-switching model for inter-trade durations in the high-frequency limit order market. *Economic Modelling*, 118:106082, 2023.

[17] Weibing Huang, Charles-Albert Lehalle, and Mathieu Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015.

[18] Hamza Bodor and Laurent Carlier. A novel approach to queue-reactive models: The importance of order sizes, 2024. arXiv preprint.

[19] A. Nittur Anantha and S. Jain. Forecasting high frequency order flow imbalance using hawkes processes. *Computational Economics*, 2025.

[20] Rama Cont, Arseniy Kukanov, and Sasha Stoikov. Price dynamics in a markovian limit order market. *SIAM Journal on Financial Mathematics*, 5(1):1–25, 2014.

[21] Weibing Huang, Charles-Albert Lehalle, and Mathieu Rosenbaum and. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015.

9 Appendix - I: Global Notation Summary

Table 3: Summary of Symbols and Their First Point of Introduction

Symbol	Description	Section
Market Structure and Events		
$p^{b,i}, p^{a,i}$	Price at level i on bid / ask side	Data / Filtration
$q^{b,i}, q^{a,i}$	Quantity at level i on bid / ask side	Data / Filtration
T	Final time in summation/indexed bounds (not event label)	Methodology
T, C	Trade and cancellation event types	Problem Description
$\mathcal{E}$	Set of event types (e.g., trades, cancels)	Problem Description
$\mathcal{I}$	Index set over event dimensions	Methodology
Feature Extraction and Window Definitions		
τ	Anchor time for evaluation window	Problem Description
h	Length of evaluation window	Problem Description
ξ	Length of forecast window	Problem Description
$[\tau - h, \tau)$	Evaluation window	Problem Description
$(\tau, \tau + \xi]$	Forecast window	Problem Description
$\tilde{r}$	Realized return over forecast window	Forecast Target
$\hat{\tilde{r}}$	Predicted return over forecast window	Experiment / Results
$\mathcal{W}$	Set of evaluation window anchors τ	Problem Description
$\mathcal{T}(w)$	Tick history over window w	Data / Preprocessing
$\mathcal{B}(w)$	LOB snapshot at start of window w	Data / Preprocessing
Filtering and Feature Construction		
$\mathscr{F}$	General filtration scheme applied to event stream	Methodology
$\mathcal{F}_q$	Quote-persistence filter	Methodology
$\mathcal{F}_e$	Event excitation filter (e.g., via Hawkes)	Methodology
$\mathcal{G}$	Optional auxiliary filter for OBI-related features	Methodology
$\mathcal{K}(w)$	Feature extraction map over window w	Methodology
Order Book Imbalance and Related Measures		
$\mathrm{OBI}_{[\tau-h,\tau)}$	Order Book Imbalance in evaluation window	Methodology
$\Delta q^{b,1}, \Delta q^{a,1}$	Changes in top-level quantities	Methodology
Hawkes Process Components		
$\lambda^i(z)$	Intensity for event type i in fused space	Methodology
μ^i	Baseline intensity for event type i	Methodology
ϕ_{ij}	Excitation kernel from j to i	Methodology
Prediction and Evaluation		
$\mathcal{S}(\mathscr{F}, \tilde{r})$	Scoring functional based on filtration and return	Methodology
$\mathcal{S}^{(\rho)}$	Correlation-based scoring functional	Methodology
$\mathcal{S}^{(\phi)}$	Excitation norm–based scoring functional (Hawkes)	Methodology
ρ	Pearson correlation coefficient	Evaluation
L	Forecast loss (e.g., absolute or squared error)	Evaluation
$\mathbb{I}\{\cdot\}$	Indicator function	Evaluation
Auxiliary Symbols and Derived Constructs		
$\mathcal{M}$	Filtered model variant or mask	Results
$\mathcal{R}$	Return regime bin identifier	Evaluation
$\mathbf{Q}, \mathbf{R}$	Feature or target vectors for OBI/returns	Methodology / Results
$\hat{Q}, \hat{R}$	Predicted values of feature or response	Evaluation

10 Appendix - II: detailed scores across filtration methods and scoring functionals.

10.1 Pearson Correlation scores across time windows

Date_Filter	1s	10s	30s	50s	80s	100s
20230102_UF	0.02067	0.00866	0.00271	0.00368	0.00428	0.00377
20230102_MTF	0.01233	0.00236	0.00070	0.00033	0.00590	0.00120
20230102_LF	0.02121	0.00945	0.00463	0.00251	0.00550	0.00332
20230102_MF	0.00092	0.00371	-0.00520	0.00088	-0.00059	-0.00151
20230113_UF	0.00353	-0.00503	0.00043	0.00071	-0.00171	-0.00334
20230113_MTF	0.01248	0.00308	-0.00048	0.00275	0.00164	-0.00182
20230113_LF	0.00379	-0.00486	0.00053	0.00116	-0.00272	-0.00374
20230113_MF	0.00123	0.00127	-0.00079	0.00247	0.00076	-0.00236
20230123_UF	0.02327	-0.00325	0.00154	-0.00044	0.00507	-0.00515
20230123_MTF	0.03022	-0.00109	0.00225	0.00184	0.00291	-0.00802
20230123_LF	0.02135	-0.00347	0.00198	-0.00104	0.00500	-0.00524
20230123_MF	0.02453	0.00075	-0.00009	0.00088	0.00035	-0.00494

Table 4: Pearson correlation magnitudes between OBI and realized returns across different filter types and lags (**Auto-Correlation included**). Filter types: UF = Unfiltered, MTF = Modification-Time Filter, LF = Lifetime Filter, MF = Modification Count Filter.

Date_Filter	1s	10s	30s	50s	80s	100s
20230102_UF	0.01192	0.00656	0.00151	0.00400	0.00250	0.00179
20230102_MTF	0.00715	0.00280	-0.00050	0.00133	0.00407	0.00009
20230102_LF	0.01219	0.00737	0.00235	0.00326	0.00327	0.00170
20230102_MF	0.00076	0.00418	-0.00330	0.00104	-0.00051	-0.00050
20230113_UF	0.00268	-0.00282	0.00026	-0.00060	-0.00097	-0.00076
20230113_MTF	0.00937	0.00262	0.00056	0.00052	0.00165	-0.00033
20230113_LF	0.00259	-0.00239	0.00059	-0.00011	-0.00155	-0.00076
20230113_MF	0.00336	0.00251	0.00008	0.00052	0.00044	-0.00086
20230123_UF	0.01692	-0.00231	0.00016	0.00062	0.00431	-0.00379
20230123_MTF	0.02145	-0.00032	0.00143	0.00262	0.00272	-0.00621
20230123_LF	0.01565	-0.00254	0.00056	0.00009	0.00410	-0.00340
20230123_MF	0.01882	0.00102	-0.00005	0.00134	0.00032	-0.00404

Table 5: Pearson correlation magnitudes between OBI and realized returns across different filter types and lags (**Auto-Correlation Removed**). Filter types: UF = Unfiltered, MTF = Modification-Time Filter, LF = Lifetime Filter, MF = Modification Count Filter.

10.2 Discretized Regime Scores across time windows

Date_Filter	CC_1s	R2_1s	CC_10s	R2_10s	CC_20s	R2_20s
20230102_UF	-14.7367	11.5045	-4.8592	9.9368	-11.9828	10.0901
20230102_MTF	-1.4220	11.4001	-0.3291	10.6411	-5.7166	9.9782
20230102_LF	-12.2667	11.4294	-3.4627	10.2400	-9.8812	10.3134
20230102_MF	5.8019	11.2123	-0.1979	10.2374	-6.9759	10.3247
20230113_UF	-2.7480	10.9939	3.3443	9.9589	0.5709	9.5452
20230113_MTF	2.0873	11.2029	-1.4341	10.0322	1.3986	9.6094
20230113_LF	-1.4729	11.4112	5.2130	10.1280	0.3083	9.7324
20230113_MF	8.7904	10.9055	-0.5019	9.9959	-4.9251	9.3581
20230123_UF	-20.6215	12.8027	5.5745	11.4413	-7.7442	11.0128
20230123_MTF	-26.5211	12.7044	4.4931	11.3200	-2.4113	11.1883
20230123_LF	-15.4226	12.7374	5.7728	11.3674	-8.5328	10.8279
20230123_MF	-11.4672	12.1819	5.1172	10.9107	-2.9086	10.5802

Table 6: Discretized regime Cross-correlation and R^2 scores between OBI and returns across filter types and time lags (**Auto-Correlation included**). Filter types: UF = Unfiltered, MTF = Modification-Time Filter, LF = Lifetime Filter, MF = Modification Count Filter, CC = Cross-Correlation

Date_Filter	CC_1s	R2_1s	CC_10s	R2_10s	CC_20s	R2_20s
20230102_UF	3.1379	7.4766	-2.6766	6.9846	-10.3337	6.9713
20230102_MTF	-3.8815	7.0980	-0.8615	6.5966	-2.5511	6.5350
20230102_LF	4.1918	7.3520	-2.0460	6.7774	-7.2939	6.8755
20230102_MF	-5.1195	7.4436	-12.5525	7.0353	-5.7030	7.2179
20230113_UF	-22.3374	7.6243	0.2531	7.1001	-3.4353	6.8736
20230113_MTF	-24.9570	7.5030	-5.4746	6.8837	-3.1096	6.8727
20230113_LF	-23.1734	7.5998	0.1417	7.0376	-3.8746	6.6639
20230113_MF	-20.0740	7.6131	-11.4961	6.8947	-3.2838	6.6064
20230123_UF	-28.2047	10.1765	3.0024	9.3359	-6.3069	8.8993
20230123_MTF	-26.9336	8.7339	-1.7586	8.0028	-3.7381	7.7381
20230123_LF	-29.3954	9.6022	4.2362	8.6465	-7.1072	8.1884
20230123_MF	-21.4077	9.9060	0.8114	8.7893	-3.9746	8.5355

Table 7: Discretized regime Cross-correlation and R^2 scores between OBI and returns across filter types and time lags (**Auto-correlation removed**). Filter types: UF = Unfiltered, MTF = Modification-Time Filter, LF = Lifetime Filter, MF = Modification Count Filter, CC = Cross-Correlation

10.3 Hawkes Kernel Norms Score

Date_Filter	SumExp
20230102_UF	8.7967
20230102_MTF	8.9752
20230102_LF	8.7149
20230102_MF	8.7292
20230113_UF	8.7664
20230113_MTF	8.8756
20230113_LF	8.7267
20230113_MF	8.8249
20230123_UF	9.2244
20230123_MTF	9.6726
20230123_LF	9.3454
20230123_MF	9.4602

Table 8: Hawkes Excitation Norms under Sum-of-Exponentials kernels. Scores reflect signal–response excitation intensity across filtered variants of OBI. Filter types: UF = Unfiltered, MTF = Modification-Time Filter, LF = Lifetime Filter, MF = Modification Count Filter.

10.4 Hawkes Kernel Norms Score (Trade OBI)

Date_Filter	SumExp
20230102_UF	10.9933
20230102_LF	15.2172
20230102_MF	13.8234
20230102_MTF	12.0679
20230113_UF	8.3639
20230113_LF	15.8630
20230113_MF	10.2922
20230113_MTF	12.7363
20230123_UF	11.5868
20230123_LF	12.0834
20230123_MF	10.8573
20230123_MTF	24.7352

Table 9: Hawkes Excitation Norms under Sum-of-Exponentials kernel, computed from OBI derived using trade events only. Scores reflect excitation intensity across filtered order streams. Filter types: UF = Unfiltered, LF = Lifetime Filter, MF = Modification Count Filter, MTF = Modification-Time Filter.