Estimation and Simulation of Rare Liquidity Events in Limit Order Books

Mohamed Tnani

June 2, 2025

Abstract

This report presents an academic project whose objective is to put into practice advanced rare event simulation techniques within the context of financial market modeling. Rather than proposing new theoretical developments, the project focuses on applying and testing existing methods—such as adaptive multilevel splitting (AMS) [4] and importance sampling [5]—to estimate the probabilities of rare but meaningful events in limit order books (LOBs). These events include market regime transitions (e.g., from bullish to bearish), sudden episodes of illiquidity, and the statistical distribution of the second bid or ask queue when the first one is depleted.

The modeling approach is based on Queue-Reactive (QR) frameworks [3, 7], which offer a simplified yet realistic representation of LOB dynamics. The goal is to assess the performance of Monte Carlo methods in these settings and to gain hands-on experience with simulation strategies that are relevant for risk management and market analysis.

Introduction

This report is part of a broader investigation into the dynamics of financial markets, with a particular focus on rare events occurring within limit order books (LOBs). Although such events are infrequent, they can have significant consequences on market stability, liquidity, and overall efficiency [2].

Modern financial market microstructure is built upon electronic systems that centralize and match buy and sell intentions. A detailed understanding of these mechanisms is essential for both market participants and regulatory authorities. In particular, modeling and predicting the statistical behavior of limit order books is crucial for designing robust trading strategies and ensuring systemic resilience in times of stress [1].

Our approach focuses on developing and analyzing Queue-Reactive (QR) models—an established class of LOB models that captures the dynamic interplay between order flows and queue positions [3, 7]. We begin with simplified

QR models and gradually enrich them to increase their realism and predictive power. However, we will not delve deeply into the simulation of complex order book dynamics, as this is not the primary objective of this study.

A key challenge lies in the estimation of rare event probabilities, where traditional Monte Carlo methods often fail due to the sheer rarity of the events of interest. To address this, we incorporate advanced simulation techniques, such as adaptive multilevel splitting (AMS) [4] and importance sampling [5], which allow for efficient and statistically sound estimation of these low-probability events.

The insights gained from this study provide new perspectives on the statistical properties of limit order books and open up promising directions for rare event modeling in high-frequency trading environments. In particular, the methodologies developed here could be useful for stress testing, market impact analysis, and the design of early warning indicators for market instability [2].

Description of the Model and Project Positioning

The Queue-Reactive (QR) model is a stochastic modeling framework designed to describe the dynamics of limit order books (LOBs) in modern electronic markets [3]. Unlike aggregate models that capture only average behaviors over time, the QR model accounts for the microscopic interactions between incoming order flows and the current state of the queue at each price level. This allows it to reflect more accurately the adaptive behavior of traders who continuously react to market conditions [7].

In electronic trading systems, buy and sell orders are placed at discrete price levels and wait in queues until execution. The QR model operates under the key assumption that the intensities of order arrivals and cancellations are not constant but rather **depend explicitly on the current sizes of the queues**. For example, traders might be more inclined to place limit orders at price levels where the queue is relatively short (thus increasing the chance of execution), or cancel their orders if the queue is too long and execution appears unlikely [3].

Mathematically, the QR model is typically formulated as a *continuous-time Markov jump process*. Let us consider a simplified LOB focusing only on the best bid and best ask. At each side of the book, we track the size of the queue over time, denoted by

$$Q(t) = (q_{\text{bid}}(t), q_{\text{ask}}(t)).$$

The state of the system evolves due to several types of stochastic events :

- **Limit order arrivals**, which increase the size of the corresponding queue, occur at a rate $\lambda^{\text{add}}(q)$

- **Market orders**, which consume liquidity and reduce queue size, arrive at a rate $\lambda^{\text{trade}}(q)$

- **Order cancellations**, which also reduce queue size, occur with intensity $\lambda^{\text{cancel}}(q)$

In all cases, the transition rates are queue-dependent ; that is,

$$\lambda^{\text{event}}(q) = f(q),$$

where f is a function that must be estimated from empirical data, often using high-frequency observations [7]. This dependence on the queue size allows the model to reflect self-regulating feedback effects inherent in real financial markets [3].

Originally proposed and formalized by Cont et al. [3], the QR modeling approach has been further validated by empirical studies including Huang et al. [7]. These contributions demonstrate that queue-reactive dynamics are essential to capturing realistic LOB behavior, particularly when analyzing short-term order flow and liquidity fluctuations.

Our analysis begins with a simplified limit order book (LOB) model consisting of a single bid queue and a single ask queue, where order arrivals, cancellations, and executions follow Markovian dynamics. While this baseline setup provides tractability, real-world LOBs exhibit feedback effects where traders actively respond to observed liquidity imbalances [2]. To capture this phenomenon, we introduce an *imbalance factor* f, defined as :

$$f = \frac{Q_{\text{bid}} - Q_{\text{ask}}}{Q_{\text{bid}} + Q_{\text{ask}}}, \tag{1}$$

where Q_{bid} and Q_{ask} denote the queue sizes at the best bid and ask levels, respectively. This metric quantifies the instantaneous liquidity skew, ranging from -1 (extreme sell-side pressure) to $+1$ (extreme buy-side dominance).

To operationalize regime classification, we partition the state space using a threshold parameter θ. The market is deemed :

- **Bullish** (Θ_+) when $f > \theta$
- **Neutral** for $|f| \leq \theta$
- **Bearish** (Θ_-) when $f < -\theta$

As established in our modeling framework, the dynamics of queue evolution must inherently depend on the instantaneous state of the order book. As summarized in Table 1, each regime associates with distinct triples $[\lambda_{\text{add}}, \lambda_{\text{cancel}}, \lambda_{\text{trade}}]$ for both bid and ask sides :

TABLE 1 – Order flow intensities by market regime

Regime	**Buy-side**	**Sell-side**
Bullish (Θ_+)	[0.8, 0.3, 0.1]	[0.4, 0.6, 0.3]
Neutral	[0.7, 0.5, 0.2]	[0.7, 0.5, 0.2]
Bearish (Θ_-)	[0.4, 0.6, 0.3]	[0.8, 0.3, 0.1]

In bullish regimes (Θ_+), buy-side limit order arrivals intensify $(\lambda_{\text{add}} = 0.8)$ while sell-side cancellations rise $(\lambda_{\text{cancel}} = 0.6)$, reflecting self-reinforcing buying momentum. Neutral regimes exhibit symmetric intensities (e.g., $\lambda_{\text{add}} = 0.7$

for both sides), mimicking balanced liquidity provision. Bearish regimes (Θ_-) invert the bullish asymmetries, with aggressive sell-side additions and buy-side cancellations.

This structure captures two essential nonlinear effects : first, liquidity providers dynamically adapt their strategies in response to queue imbalances, and second, even modest imbalances may induce regime transitions that reinforce existing trends. This behavior aligns with empirical findings regarding liquidity evaporation during periods of market stress [2].

An example of simulated order book trajectories under the proposed model is presented in Figure 1, with initial queue depths $Q_{\text{bid}}(0) = Q_{\text{ask}}(0) = 50$.

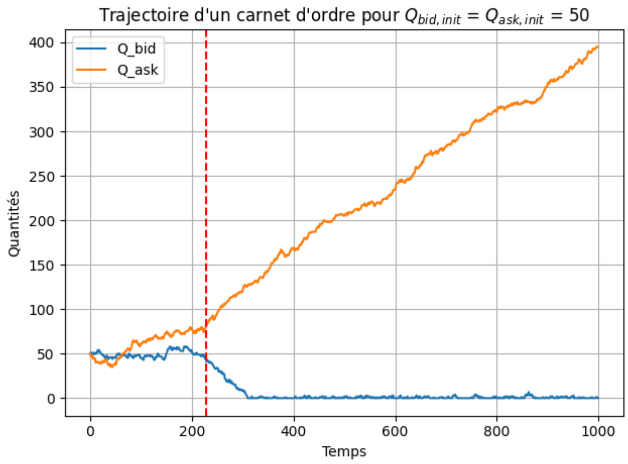

FIGURE 1 – Order book dynamics simulation with $Q_{\text{bid,init}} = Q_{\text{ask,init}} = 50$. The red vertical line indicates the transition time from neutral to Θ_+ regime

Imbalnce Impact on Simulations

To study how initial imbalances affect order book dynamics, we focused on two key scenarios : when the ask queue completely depletes ($A_{\text{ask}} = \{Q_{\text{ask}}(\infty) = 0\}$) and when the bid queue depletes ($A_{\text{bid}} = \{Q_{\text{bid}}(\infty) = 0\}$). We ran extended simulations, with each trial continuing until $T = 10,000$ econds to ensure sufficient time for the queues to potentially empty.

Using naive Monter Carlo Simualtions, we estimated two critical probabilities : the chance of ask queue depletion when starting from a strongly bullish position ($f_{\text{init}} = +\theta$), and the probability of bid queue depletion when beginning from a strongly bearish position ($f_{\text{init}} = -\theta$). After running numerous simulations, the results showed remarkably consistent patterns.

When the order book began in a bullish state ($f = +\theta$), we found the ask queue eventually emptied in 94.12% of cases, with a tight 95% confidence interval of $\pm 0.86\%$. Conversely, when starting from a bearish position ($f = -\theta$), the bid queue depleted in 95.08% of simulations, with an even narrower confidence interval of $\pm 0.56\%$.

These near-95% probabilities demonstrate how initial imbalances tend to persist and amplify over time. Once the order book enters either the strongly bullish or bearish regime, it overwhelmingly continues in that direction until one side's liquidity is exhausted. Second, the slight difference in probabilities and confidence intervals between bullish and bearish scenarios may reflect subtle asymmetries in how traders react to buying versus selling pressure.

Notations

To formalize our analysis of regime transitions, we define the following key temporal metrics :

Notation	Definition
τ_+	First transition time to Θ_+ state (neutral $\to$ bullish)
τ_-	First transition time to Θ_- state (neutral $\to$ bearish)
$\min(\tau_+, \tau_-)$	Exit time from neutral state ($f_{\text{init}} \in [-\theta, +\theta]$)

Table 2 – Definition of market regime transition times.

1 First Transition from Neutral Initial States

1.1 Identifying the rare event :

The probability $\mathbb{P}(\tau_+ > \tau_- \mid f_{\text{init}} \in [-\theta, \theta])$ quantifies whether an order book starting in neutral state ($|f_{\text{init}}| \leq \theta$) will transition first to the bearish regime Θ_- rather than the bullish regime Θ_+. This metric reveals inherent asymmetries in liquidity dynamics - a value significantly above 0.5 would indicate systemic bias toward bearish transitions even from equilibrium conditions, while values near 0.5 suggest symmetric regime switching.

To estimate this probability, we simulate order book trajectories with initial neutral imbalances, tracking the first exit time $\min(\tau_+, \tau_-)$. The cutoff time $T_{\max}$ is chosen sufficiently large (typically 10,000 seconds) to ensure most trajectories experience at least one transition.

To systematically analyze regime transitions from neutral initial conditions, we varied the initial bid queue size $Q_{\text{bid init}}$ to generate imbalance values f_{init} spanning the neutral range $[-\theta, +\theta]$.

Figure 2 displays the resulting transition probabilities $\mathbb{P}(\tau_+ > \tau_- \mid f_{init})$ across different initial imbalance values and threshold configurations.

Our analysis of transition probabilities reveals that $P(\tau_+ < \tau_- \mid f_{\text{init}}, \theta)$ exhibits two key properties. First, the probability increases with the initial imbalance f_{init} – the closer the order book starts to the bearish boundary ($-\theta$), the less likely it is to transition to the bullish regime (Θ_+) before becoming fully bearish. Second, for a fixed neighborhood width ϵ, this probability also

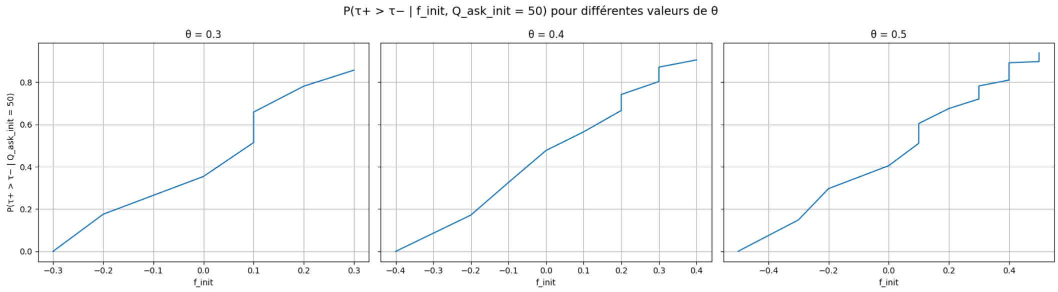

FIGURE 2 – $\mathbb{P}(\tau_+ > \tau_- | f_{init}, Q_{ask\_init} = 50)$ for $f_{init} \in [-\theta, +\theta]$. The simulation parameters were set to $T_{max} = 1.10^4$, $N_{\text{sim}} = 2.10^3$, $Q_{\text{ask init}} = 50$, and $\theta \in \{0.3, 0.4, 0.5\}$.

increases with the threshold θ itself, as larger θ values require the order book to traverse a greater "distance" from near-bearish conditions $(-\theta + \epsilon)$ to reach the bullish regime.

These observations led us to identify :

$$B_{\theta,\epsilon} = \{\tau_+ < \tau_- \mid f_{\text{init}} = -\theta + \epsilon\} \tag{2}$$

as a particularly interesting rare event. This represents cases where the market, despite beginning at the very edge of bearish territory (just ϵ away from $-\theta$), manages to reverse course and enter a bullish regime first. As shown in Figure 2, standard Monte Carlo methods fail to reliably estimate these probabilities, especially for larger θ values like 0.5 – the events simply occur too infrequently in naive simulations.

The computational challenges are most severe when θ is large (0.5 vs 0.3), making the required imbalance reversal more extreme, and ϵ is small, meaning we start closer to the bearish threshold. This motivates our use of an advanced simulation technique (AMS) to accurately quantify these rare but economically significant events.

1.2 AMS Splitting Algorihtm on $\mathbb{P}(B_{\theta,\epsilon}|\theta, \epsilon)$

To estimate the transition probability $\mathbb{P}(B_{\theta,\epsilon}) = \mathbb{P}(\tau_+ < \tau_- \mid f_{\text{init}} = -\theta+\epsilon)$, we employ Adaptive Multilevel Splitting (AMS) [4]. The method proceeds through the following steps.

First, we discretize the imbalance range $[-\theta+\epsilon, \theta]$ into equally spaced levels with step size $\Delta f = 0.05$, defining thresholds :

$$f_k = -\theta + \epsilon + k\Delta f \quad \text{for} \quad k = 0, \dots, K$$

where K is the smallest integer such that $f_K \geq \theta$.

The AMS algorithm begins simulation from initial trajectories starting at $f_0 = -\theta+\epsilon$. At each iteration k, the order book dynamics are simulated forward in time until every trajectory either successfully crosses the next threshold f_{k+1} or fails by entering the bearish regime Θ_-. The survival probability $\hat{p}_k$ for that level is then calculated as the fraction of trajectories that successfully reached f_{k+1} out of the total number simulated, expressed mathematically as :

$$\hat{p}_k = \frac{N_{\text{surviving}}}{N_{\text{total}}}$$

These surviving trajectories are then randomly resampled with replacement to serve as the starting population for the next iteration targeting threshold f_{k+2}. This resampling ensures sufficient trajectory diversity while progressively focusing computational effort on the increasingly rare paths that continue advancing toward the bullish regime Θ_+. The process repeats until either the target threshold θ is crossed or all trajectories fail.

The target probability is then obtained as the product :

$$\mathbb{P}(B_{\theta,\epsilon}) \approx \prod_{k=0}^{K-1} \hat{p}_k \tag{3}$$

This approach efficiently handles the rare event scenario, where standard Monte Carlo methods fail to provide reliable estimates (see Figure 2).

1.3 Biais Evaluation

To evaluate the robustness of our AMS algorithm, we conduct $N = 20$ independent simulations for each parameter combination (θ, ϵ), generating probability estimates $\{p_1, p_2, \ldots, p_N\}$.

We construct a global estimator as the global mean of estimators, and, assuming $\hat{p}$ is unbiased, we construct a 95% confidence intervals :

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{N} p_i \;\pm\; q_{0.975} \cdot \frac{s}{\sqrt{N}} \tag{4}$$

where $q_{0.975} \approx 1.96$ is the 97.5% quantile of the standard normal distribution and s is the sample standard deviation.

We observe, as shown in Table 1.3, that $\mathbb{P}(B_{\theta,\epsilon} \mid \theta, \epsilon)$ is generally higher when θ is smaller. This is due to the fact that there are fewer levels to cross in the AMS in the first case.

We also note that the dispersion of the estimators remains, in most cases, within the confidence intervals. This confirms that our estimator is unbiased, which implies the robustness of our algorithm. See [4] for more details.

2 Distribution of Transition Time $\min(\tau_+, \tau_-)$

This section analyzes the **transition time** $\min(\tau_+, \tau_-)$, defined as the time required for the order book to reach a non-neutral state from a given initial

θ	**Estimated** $\mathbb{P}(B_{\theta,\epsilon})$
0.3	$1.10^{-4} \pm 2.10^{-5}$
0.4	$5.10^{-5} \pm 1.10^{-5}$

TABLE 3 – Estimated $\mathbb{P}(B_{\theta,\epsilon})$ using AMS Splitting, with the following parameters : $T_{\max} = 10^4$, $\Delta f = 0.05$, $\epsilon = 0.01$, $\theta \in [0.3, 0.4]$.

condition. This metric serves as a proxy for **market stability** : a lower expected transition time $\mathbb{E}[\min(\tau_+, \tau_-)]$ indicates faster convergence to a new equilibrium, implying greater price stability.

We estimate the distribution $\mathcal{L}(\min(\tau_+, \tau_-))$ using a naïve Monte Carlo simulation. The results, shown in Figure 3, reveal two key trends.

First, the distribution decays rapidly across all tested configurations, irrespective of the threshold θ or initial imbalance f_{init}. This suggests that most order book trajectories reach a non-neutral state quickly, reflecting highly dynamic market conditions.

Second, while the plotted distributions are truncated at $T = 3.10^3$ for visual clarity, empirical observations confirm that a small fraction of simulations exhibit significantly longer transition times. These outliers correspond to scenarios where the order book maintains a temporary equilibrium due to balanced order flow.

2.1 Impact of the initial depth on $\min(\tau_+, \tau_-)$

Building on the transition time analysis, we now investigate how initial order book quantities influence market stability by estimating $\mathbb{E}[\min(\tau_+, \tau_-) \mid Q_{\text{bid,init}} = Q_{\text{ask,init}} = q, \theta]$. This conditional expectation measures the average transition time for perfectly neutral markets ($f_{\text{init}} = 0$) across varying initial depths q.

The results shown in figure a clear positive relationship between initial depth q and expected transition time. For fixed θ, $\mathbb{E}[\min(\tau_+, \tau_-)]$ increases monotonically with q, indicating that deeper order books (greater liquidity on both sides) exhibit slower transitions to non-neutral states. Following [2], the impact of a unit market order $\Delta f \approx 2/q^2$ for $Q_{\text{ask}} \approx Q_{\text{bid}} \approx q$, confirming the inverse-square law of depth impact observed in our simulations.

2.2 Estimating the Persistence Probability $\mathbb{P}(\min(\tau_+, \tau_-) > T \mid \theta)$

Building on our transition time analysis, we now investigate the persistence probability of $R_{\theta,T} = \{\min(\tau_+, \tau_-) > T \mid \theta\}$. This quantity measures the likelihood that the order book remains in the neutral state beyond a substantial time horizon T, indicating prolonged market stability without transitions to bullish or bearish regimes.

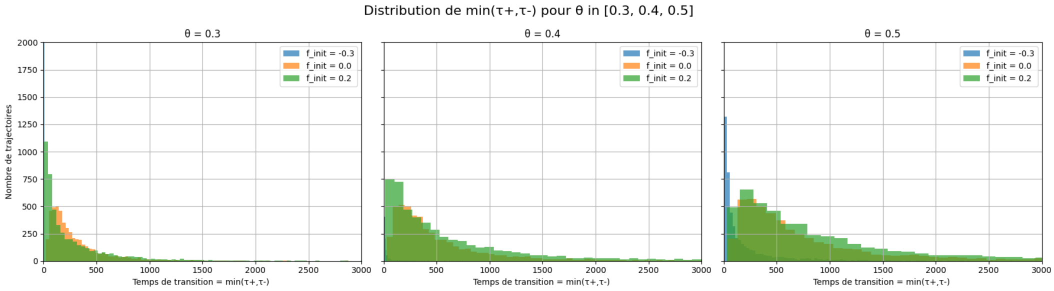

FIGURE 3 – Distribution of $\min(\tau_+, \tau_-)$ for thresholds $\theta \in \{0.3, 0.4, 0.5\}$. $T_{\text{max}} = 2.10^4$, $N_{\text{sim}} = 2,5.10^4$ Monte Carlo runs, and initial ask quantities $Q_{\text{ask,init}} \in \{25, 50, 75\}$ to span $f_{\text{init}} \in \{-0.3, 0, 0.2\}$..

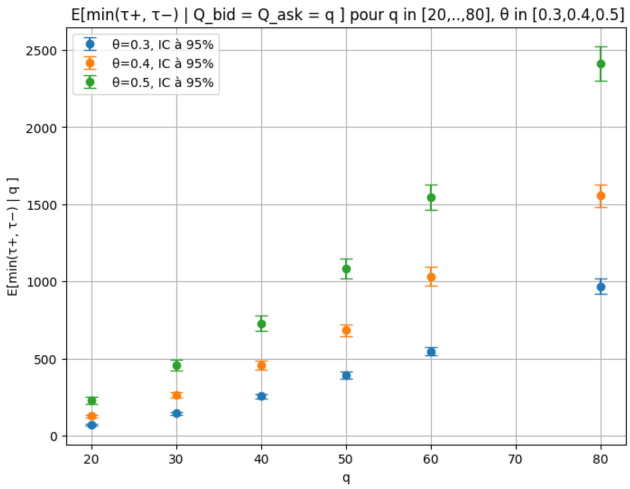

FIGURE 4 – $\mathbb{E}[\min(\tau_+, \tau_-) \mid Q_{\text{bid}} = Q_{\text{ask}} = q]$ for $q \in [20, 80]$ and $\theta \in \{0.3, 0.4, 0.5\}$. $T_{\text{max}} = 2.10^4$, $N_{\text{sim}} = 2,5.10^4$ independent runs Error bars represent 95% confidence intervals.

To estimate this rare event probability efficiently, we employ the AMS Algorithm introduced in 1.2. We discretize the time interval $[0, T]$ into steps of size Δt. At each stage k, we compute the conditional survival probability to time $k\Delta t$ given survival to $(k-1)\Delta t$. Trajectories that successfully reach each successive threshold are resampled, while non-surviving paths are discarded.

Similarly to (3), the target probability is then obtained as the product :

$$\mathbb{P}(R_{\theta,T}) = \prod_{k=1}^{T/\Delta t} \mathbb{P}(\text{survive to } k\Delta t \mid \text{survived to } (k-1)\Delta t),$$

Following the methodology established in Section 1.3, we compute a global mean estimator for $\mathbb{P}(R_{\theta,T})$ through N independent simulations. The confidence intervals are derived using the same central limit theorem approach demonstrated previously, as formalized in Equation (4). Table 4 shows the results for different (θ, T) values.

T	$\theta = 0.3$	$\theta = 0.4$
3.10^3	22 ± 6	202 ± 10
4.10^3	12 ± 4	135 ± 10
5.10^3	8 ± 3	70 ± 7

TABLE 4 – Estimated $\mathbb{P}(R_{\theta,T})$ using AMS Splitting. Simulation parameters : $Q_{\text{bid,init}} = Q_{\text{ask,init}} = 50$, $\Delta t = 3.10^2$, $T \in [3, 4, 5].10^2$, $\theta \in [0.3, 0.4]$. Results are scaled by 10^4.

Prolonged neutral states ($R_{\theta,T} \gg 0$) suggest either strong liquidity provision maintaining equilibrium or insufficient order flow to trigger regime transitions. Our single-queue model captures this as temporary balance between bid/ask order arrivals and cancellations. The AMS estimator's variance reduction enables precise quantification even for probabilities below 10^{-3}, revealing how market depth θ affects stability timescales.

3 Two-Level Order Book Model with Bid/Ask Queues

The model extends the basic order book structure by introducing a second-level bid/ask queue that evolves according to independent Poisson processes. The net change in the second-level queue is given by :

$$\Delta Q^{(2)}_{\text{bid/ask}} = X - Y \quad \text{where} \quad X \sim \text{Poisson}(\lambda^{(2)}_{\text{add}}), \quad Y \sim \text{Poisson}(\lambda^{(2)}_{\text{cancel}}) \quad (5)$$

The intensity parameters $\lambda^{(2)}_{\text{add}}$ and $\lambda^{(2)}_{\text{cancel}}$ are dynamically adjusted based on the relative liquidity ratio :

$$r_{\text{bid/ask}}^{(2)} = \frac{Q_{\text{bid/ask}}^{(2)}}{Q_{\text{bid/ask}}^{(1)} + Q_{\text{bid/ask}}^{(2)}} \tag{6}$$

The intensity parameters follow a threshold-based adjustment rule :

$$\lambda_{\text{add}}^{(2)}, \lambda_{\text{cancel}}^{(2)} = \begin{cases} (\lambda_{\text{add,high}}^{(2)}, \lambda_{\text{cancel,high}}^{(2)}) & \text{if } r_{\text{bid/ask}}^{(2)} > s \\ (\lambda_{\text{add,low}}^{(2)}, \lambda_{\text{cancel,low}}^{(2)}) & \text{if } r_{\text{bid/ask}}^{(2)} \leq s \end{cases}$$

where s represents the liquidity threshold. This formulation captures the market's tendency to adjust order flow dynamics based on the relative distribution of liquidity between price levels.

For the simulation studies, we used the parameter values shown in Table 3.

Parameter	$r_{\text{bid/ask}}^{(2)} > 0.2$	$r_{\text{bid/ask}}^{(2)} \leq 0.2$
$\lambda_{\text{add}}^{(2)}$	0.6	0.8
$\lambda_{\text{cancel}}^{(2)}$	0.5	0.4
Threshold s	0.2	

TABLE 5 – Second Queues' Simulation Parameters

3.1 Distribution of $Q_{2,\text{bid}}$ When $Q_{1,\text{bid}} = 0$

We examine the distribution of the second-level bid queue when the first-level depletes :

$$Q_{2,\text{bid}}(T^+) \,\big|\, Q_{1,\text{bid,init}}, Q_{1,\text{ask,init}}, \theta \tag{7}$$

where $T^+ = \min\{t \geq 0 : Q_{1,\text{bid}}(t) = 0\}$ and we consider only trajectories where $T^+ < T^-$ (with T^- being the analogous depletion time for the ask side).

Analyzing this distribution is crucial for two main reasons. First, it reveals residual liquidity at the second level when the top bid vanishes, crucial during market stress [2]. This helps assess the market's ability to absorb large orders without catastrophic price moves.

Moreover, the shape of this distribution directly quantifies slippage risk [6]. When $Q_{2,\text{bid}}$ is typically small at T^+, the market becomes vulnerable to volatility spikes as liquidity vanishes across multiple levels.

We estimate the distribution of (7) using Naive Monter Carlo simulations, the results are shown in Figure 5.

We observe that for a fixed initial bid queue depth $Q_{1,\text{bid}}^{\text{init}}$, the distribution tends to shift rightward as θ increases. This indicates more trajectories where the apparent order book imbalance favors the bid side.

Two key mechanisms explain this phenomenon. First, the expected transition time $\mathbb{E}[\tau^+ \mid Q_{1,\text{bid}}^{\text{init}}]$ increases with θ. Reaching the threshold τ^+ becomes

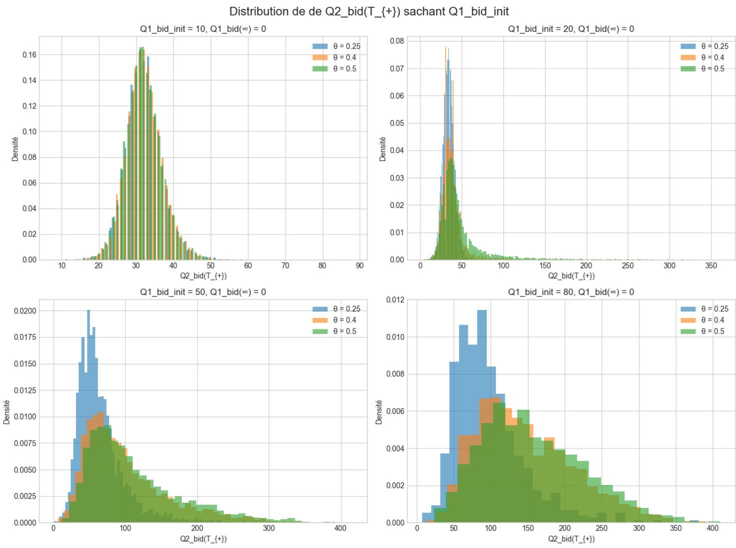

FIGURE 5 – Distribution of $Q_{2,\text{bid}}$ at depletion time T^+. Simulation parameters : $N_{\text{sim}} = 2, 5.10^4$, $T_{\max} = 10^4$, $Q_{1,\text{ask,init}} = 50$, $Q_{1,\text{bid,init}} \in \{10, 20, 50, 80\}$, $\theta \in \{0.25, 0.4, 0.5\}$.

statistically harder at higher θ values, requiring more time for the imbalance to achieve the critical level.

Second, the second-level bid queue $Q_{2,\text{bid}}$ exhibits positive drift due to the net order flow :

$$\mathbb{E}[\Delta Q_{2,\text{bid}}] = \lambda_{\text{add}} - \lambda_{\text{cancel}} > 0 \tag{8}$$

This ensures $Q_{2,\text{bid}}$ grows progressively during the extended transition period before $Q_{1,\text{bid}}$ depletion. The longer time window (T^+) at higher θ therefore allows more accumulation of depth at the second price level.

3.2 Estimating Tail Probabilities of Residual Liquidity

We analyze the probability that the second-level bid queue exceeds a specified threshold when the first-level depletes :

$$\mathbb{P}\left(Q_{2,\text{bid}}(T^+) > Q_{\text{threshold}} \mid Q_{1,\text{bid}}^{\text{init}}, Q_{1,\text{ask}}^{\text{init}}, \theta\right) \tag{9}$$

where $T^+ = \inf\{t \geq 0 : Q_{1,\text{bid}}(t) = 0\}$. The positive net order flow $\mathbb{E}[\Delta Q_{2,\text{bid}}] = \lambda_{\text{add}} - \lambda_{\text{cancel}} > 0$ suggests this probability increases with the arrival rate.

A natural approach to estimate (9) is to increase the expected queue growth (8) to make large terminal values $Q_{2,\text{bid}}(T^+)$ occur more frequently. This can be achieved through *importance sampling* [5] by modifying the original Poisson rates $(\lambda_{2,\text{add}}, \lambda_{2,\text{cancel}})$ to new rates $(\lambda'_{2,\text{add}}, \lambda'_{2,\text{cancel}})$.

Thus, under this transformed measure, the queue dynamics become :

$$Q_{2,\text{bid}}(t) = \underbrace{k_1(t)}_{\text{PP}(\lambda'_{2,\text{add}})} - \underbrace{k_2(t)}_{\text{PP}(\lambda'_{2,\text{cancel}})} \tag{10}$$

where $k_1(t)$,$k_2(t)$ count respectfully order additions and order cancellations. The likelihood ratio, for this measure change is given by :

$$\frac{d\mathbb{P}}{d\mathbb{P}'} = e^{(\lambda_{2,\text{add}}-\lambda'_{2,\text{add}})T - f_{\text{add}}k_1(T)} \cdot e^{(\lambda_{2,\text{cancel}}-\lambda'_{2,\text{cancel}})T - f_{\text{cancel}}k_2(T)} \tag{11}$$

where $f_{\text{add}} = \log(\lambda'_{2,\text{add}}/\lambda_{2,\text{add}})$ and $f_{\text{cancel}} = \log(\lambda'_{2,\text{cancel}}/\lambda_{2,\text{cancel}})$.

In our problem, we focus on the indicator function $g(K_{1,T}, K_{2,T}) = \mathbf{1}_{\{k_1(T)-k_2(T)>Q_{\text{threshold}}\}}$, which leads to the following estimator :

$$\mathbb{E}[g] = \mathbb{E}'\left[g \cdot e^{(\lambda_{2,\text{add}}-\lambda'_{2,\text{add}})T - f_{\text{add}}k_1(T)} \cdot e^{(\lambda_{2,\text{cancel}}-\lambda'_{2,\text{cancel}})T - f_{\text{cancel}}k_2(T)}\right] \tag{12}$$

3.3 Limitations and Proposed Solutions

While theoretically sound, our importance sampling approach faces two implementation challenges in the order book context :

1. **Random Stopping Time Problem :** The algorithm assumes a fixed time horizon T, whereas in reality, the first-level bid depletion time T^+ is stochastic. This creates a fundamental inconsistency, as our measure change theory requires deterministic time boundaries. The random nature of T^+ means we cannot directly apply standard importance sampling results derived for fixed-time scenarios.
2. **Dynamic Coupling Issue :** Modifying $\lambda_{\text{add},2}$ inadvertently affects the entire system dynamics because the liquidity ratio $r_2 = Q_{2,\text{bid}}/(Q_{1,\text{bid}} + Q_{2,\text{bid}})$ governs the state-dependent intensity parameters. This creates a feedback loop : changing the arrival rate alters the queue ratio, which in turn modifies the very intensities we attempted to control. Consequently, we lose the ability to predict the exact distributional impact of our parameter adjustments.

To address these challenges, we implement two key modifications.

-We first treat $\lambda_{\text{add},2}$ as a constant parameter, independent of r_2. This simplification breaks the feedback loop, making $k_1(t)$ a standard Poisson process. While this approximation requires validation, empirical studies of limit order books [2] suggest that addition rates at deeper price levels exhibit weaker state-dependence than cancellation rates.

-For the stopping time, we estimate first $\mathbb{E}[\mathbf{1}_{\{Q_{2,\text{bid}}(t)>Q_{\text{threshold}}\}} \mid T^+ = t]$ for fixed t using importance sampling, then integrate over the distribution of $T^+ \mid Q^{\text{init}}_{1,\text{bid}}, Q^{\text{init}}_{1,\text{ask}}, \theta$.

The estimation decomposes,then, naturally via the law of total probability :

$$p_{\theta,Q_{th}} = \int_0^\infty \mathbb{P}(Q_{2,\text{bid}}(t) > Q_{\text{th}} \mid T^+ = t)\, dF_\theta(t) \tag{13}$$

The empirical distribution function $\hat{F}_\theta(t)$ of $T^+ \mid Q_{1,\text{bid}}^{\text{init}}, Q_{1,\text{ask}}^{\text{init}}, \theta$ is constructed through Monte Carlo simulation of the order book dynamics.

For the conditional probabilities, we employ the algorithm presented in 3.2, since the simulation parameters of $Q_{2,bid}(t)$ have been uncorrelated from the rest of the model.

3.4 Computational Implementation and Results

To accelerate the Monte Carlo simulation, we simulate T^+ for each path and terminate the simulation as soon as T^+ reaches the predefined threshold $T_{\max}$. This avoids spending computational time on trajectories where T^+ takes extremely large.

Consequently, we estimate the distribution of $T^+ \mid$ on $T^+ < T_{\max}$. The choice of $T_{\max}$ will be discussed in the following subsection.

Under this truncation, the estimator for $\hat{p}(\theta)$ becomes :

$$\hat{p}_{\theta,Q_{th},T_{max}} = \sum_{t=1}^{T_{\max}} \hat{q}_t \cdot \Delta\hat{F}_\theta(t), \tag{14}$$

where $\hat{q}_t$ estimates the conditional probability at time t and $\Delta\hat{F}_\theta(t)$ represents the empirical probability mass of T^+.

With formula 14, we cannot construct an explicit confidence interval. Therefore, we propose to repeat the estimation procedure N times, and to take as the final estimator the mean of these estimates. An asymptotic confidence interval can then be constructed in the same way as in 4. Final results are shown in Figure 6.

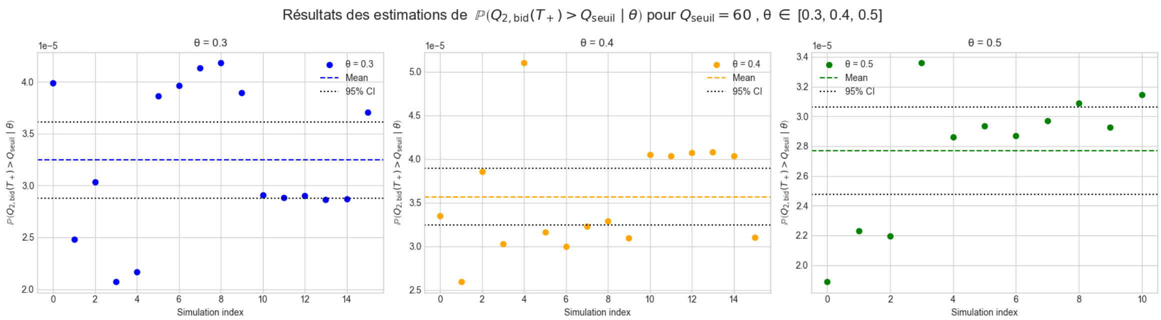

FIGURE 6 – Estimation results for $\hat{p}_{\theta,Q_{th},T_{max}}$ with $Q_{\text{th}} = 60$, $Q_{1,\text{ask}}^{\text{init}} = 60$, $Q_{1,\text{bid}}^{\text{init}} = 10$, $\theta \in \{0.3, 0.4, 0.5\}$, $N = 15$ and $T_{max} = 200$.

We can observe in Figure 6 that the estimator presented in (6) is not robust, as it exhibits a significant number of outliers in the estimated values.

This lack of robustness can be explained by the fact that the estimator is defined as a sum of several intermediate estimators. Since each intermediate estimator carries its own variability, their combination can amplify extreme values, thereby increasing the dispersion of the final estimator and making it more sensitive to fluctuations in the individual components. However, we can still conclude on the approximate order of magnitude of $p_{\theta,Q_{\text{th}},T_{\max}}$, which is around 10^{-5}.

3.5 Choice of T_{max}

Since the order of magnitude of $p_{\theta,Q_{\text{th}},T_{\max}}$ is around 10^{-5}, we can validate the sum truncation in (14) if $\mathbb{P}\big(T > T_+ \,\big|\, Q_{1,\text{bid}}^{\text{init}}, Q_{1,\text{ask}}^{\text{init}}, \theta\big)$ is at least smaller than 10^{-6}. Otherwise, the truncation is not valid. Given that the target order of magnitude is small, we propose to estimate this probability using the AMS splitting method. The procedure is similar to the estimation of $R_{\theta,T}$ presented in (2.2).

The final estimated probabilities are reported in Table 3.5.

	$\theta = 0.3$	$\theta = 0.4$	$\theta = 0.5$
$\hat{p}$	1.0×10^{-9}	0.18×10^{-8}	0.95×10^{-8}
$\text{IC}_{95\%}$	$[0.35,\ 1.55] \times 10^{-9}$	$[0.045,\ 0.45] \times 10^{-8}$	$[0.35,\ 2.35] \times 10^{-8}$

Table 6 – Estimated $\mathbb{P}(T_+ > T_{max} = 200)$ using AMS Splitting, for $Q_{1,\text{ask}}^{\text{init}} = 60$, $Q_{1,\text{bid}}^{\text{init}} = 10$. AMS parameters : $N = 500$ trajectories per phase, $\Delta t = 20$ (threshold interval for splitting).

Based on the results of the last simulation, we observe that $\mathbb{P}(T_+ > 200 \mid \theta)$ is on the order of 10^{-9} to 10^{-8} for the given initial order book parameters. This probability is extremely small — in fact, negligible — compared to $\mathbb{P}\big(Q_{2,\text{bid}}(T_+) > Q_{\text{threshold}} \mid \theta, T_+ \leq 200\big)$. We therefore conclude that the truncation used in equation (14) is valid, which implies that the final magnitude of $\hat{p}_{\theta,Q_{\text{th}}}$ is approximately 10^{-5}.

Conclusion

This project has provided an in-depth exploration of rare event modeling in financial limit order books through a progressive and rigorous methodological framework. Inspired by the seminal works of [3] and [7], we developed Queue-Reactive (QR) models of increasing complexity to capture the feedback effects between instantaneous order book imbalance and the intensities of order arrivals.

A central methodological contribution of this work was the implementation of the Adaptive Multilevel Splitting (AMS) method [4] to efficiently estimate probabilities of extremely rare events, with magnitudes ranging from 10^{-5} to

10^{-8}. Compared to naïve Monte Carlo simulations, AMS delivered both statistical robustness and substantial computational gains, confirming its relevance for high-dimensional rare event problems. These rare events include transitions from quasi-bearish to bullish market states, which—despite their low probability—can have significant economic impact.

By extending the model to track the quantity available at the second bid level at the moment the best bid is depleted, we capture an important dimension of market resilience. It also provides a direct assessment of slippage risk [6] : when liquidity at the second level is typically low at the moment of top-level depletion, markets are more vulnerable to sharp price jumps as liquidity evaporates across multiple price tiers. Our combined approach using AMS Spliting and Importance Sampling to estimate the tail distribution provides an effective framework for risk assessment, offering crucial insights for risk management applications.

Références

[1] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Market liquidity and trading activity. *Journal of Finance*, 2001.

[2] Rama Cont and Lakshithe Wagalath. Running for the exit : Distressed selling and endogenous correlation in financial markets. *Mathematical Finance*, 2014.

[3] Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations Research*, 2010.

[4] Frédéric Cérou and Arnaud Guyader. Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 2007.

[5] Paul Glasserman, Philip Heidelberger, and Perwez Shahabuddin. Importance sampling and stratification for value-at-risk. *Computational Finance*, 1999.

[6] Martin D Gould, Mason A Porter, Sam Williams, Mark McDonald, Daniel J Fenn, and Sam Howison. Limit order book liquidity. *Quantitative Finance*, 2013.

[7] Weibing Huang, Charles-Albert Lehalle, and Mathieu Rosenbaum. Simulating and analyzing order book data : The queue-reactive model. *Journal of the American Statistical Association*, 2015.