A Novel Approach to Queue-Reactive Models: The Importance of Order Sizes

Hamza Bodor[a,b], Laurent Carlier[b]

[a]*Université Paris 1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne, 106 Boulevard de l'Hôpital, Paris Cedex 13, 75642, France*
[b]*BNP Paribas Corporate and Institutional Banking, Global Markets Data & Artificial Intelligence Lab, 20 boulevard des Italiens, Paris, 75009, France*

Abstract

In this study, we delve into the application and extension of the queue-reactive model for the simulation of limit order books. Our approach emphasizes the importance of order sizes, in conjunction with their type and arrival rate, by integrating the current state of the order book to determine, not only the intensity of order arrivals and their type, but also their sizes. These extensions generate simulated markets that are in line with numerous stylized facts of the market. Our empirical calibration, using futures on German bonds, reveals that the extended queue-reactive model significantly improves the description of order flow properties and the shape of queue distributions. Moreover, our findings demonstrate that the extended model produces simulated markets with a volatility comparable to historical real data, utilizing only endogenous information from the limit order book. This research underscores the potential of the queue-reactive model and its extensions in accurately simulating market dynamics and providing valuable insights into the complex nature of limit order book modeling.

Keywords: Limit order book, Market simulators, Stylized facts, Market microstructure

Email addresses: `bodor.hamza@gmail.com` (Hamza Bodor), `laurent.carlier@bnpparibas.com` (Laurent Carlier)

December 21, 2023

1. Introduction

Limit order book (LOB) simulation is a critical area of study with far-reaching implications in the field of finance. It serves as the backbone for understanding and predicting market dynamics, thereby aiding in the development of effective trading strategies (Cartea et al., 2015). The primary utility of LOB simulation is manifested in its application to the development and assessment of algorithmic trading strategies. By simulating various market scenarios and order book dynamics, LOB models provide crucial insights that enable traders to formulate strategies that are responsive to different market conditions. This not only aids in understanding emerging market trends but also plays a significant role in enhancing the decision-making processes in algorithmic trading (Cohen and Szpruch, 2012; Balch et al., 2019). Additionally, it plays a pivotal role in risk management, where understanding the LOB can help mitigate potential losses due to market volatility (Byrd et al., 2019). The goal of LOB simulation is to accurately replicate the complex dynamics of financial markets. This involves capturing the intricate interactions between different market participants, the impact of external events on market behavior, and the inherent uncertainty and randomness in market movements. However, achieving a high degree of accuracy in LOB simulation is a complex task. The intrinsic complexity of financial markets, coupled with the vast amount of data involved and the need for high computational efficiency, makes this a challenging problem.

Alternatively, agent-based methods model the limit order book as a system of interacting agents, each following a set of rules or strategies. These methods can capture the heterogeneity of market participants and the strategic interactions between them. ABIDES Byrd et al. (2019), an interactive agent-based discrete event simulation environment, has encouraged multiple works on agent-based limit order book generation (Vyetrenko et al., 2020; Amrouni et al., 2021). More recently, Lussange et al. (2021) propose an agent composed of two parts: one part predicts the asset price based on the state of the order book, and the second component uses this prediction to decide on the action to take. They demonstrate that a configuration with agents of this type, along with Zero Intelligence (ZI) agents, allows for the production of a simulated market faithful to multiple stylized facts of the equity market.

Recently, machine learning techniques, particularly Generative Adversarial Networks (GANs), have also been applied to limit order book simulation (Sirignano and Cont, 2019; Zhang et al., 2019; Coletta et al., 2023). These methods can learn complex patterns from data and generate realistic order book dynamics. They have shown promising results and good capability of taking into account different aspects of the limit order book.

2. Methodology

2.1. Queue-Reactive Model

The Queue-Reactive (QR) model of Huang et al. (2015) has gained significant attention from numerous researchers due to its simple yet effective simulation framework for the limit order book. This model verifies multiple stylized facts and addresses various issues that other basic simulators may encounter (Mariotti et al., 2023).

In [Huang et al. (2015)](#), the limit order book is viewed as a $2K$-dimensional vector, where K represents the depth of the LOB. The reference price, often[1] the mid-price, divides the LOB into two halves: the bid side and the ask side. Each side is populated with an aggregation of limit orders, represented by $Q_{\pm i}$, where i denotes the distance from the reference price in terms of ticks.

The quantities at each limit are subject to changes due to three types of orders: *Limit* (L) orders, which increase the size of the queue; *Cancel* (C) orders, which remove an existing order; and *Market* (M) orders, which immediately match with an existing limit quantity, resulting in trades (hence, this type may also be referred to as *Trade* orders).

In the queue-reactive model, the arrival of order flow at a given price level is modeled as a heterogeneous Poisson process. The intensity of this process depends solely on the current state of the order book, particularly the queue sizes. The intensities of limit, cancel, and market order arrivals are denoted by λ^L, λ^C, and λ^M respectively, and these intensities are functions of the queue size at each limit.
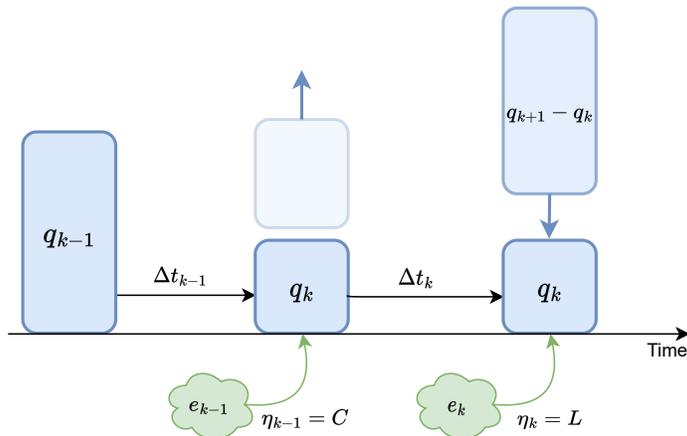

Figure 1: Example of queue updates scenario.

Figure 1 shows how the queue size of a given level is updated. where event $k-1$ decreases the size of the queue from q_{k-1} to q_k after Δt_{k-1}. Similarly, the size of the queue increases to q_{k+1} after Δt_k when event k occurs.

While this explains how the queue sizes evolve under the queue-reactive model, it's also important to understand how the price changes under this model. In fact, Figure 2 displays a scenario of price change under the QR model where this occurs when the best bid or the best ask price is totally consumed (either after a cancel or a market order). When this happens and the spread value is an even multiplier of the tick size, the reference price p_{ref} will now have two possible values: $p_{mid} + 0.5$ tick or $p_{mid} - 0.5$ tick. The model postulates then that with probability θ, the reference price changes value, allowing the

---

[1]The reference price typically matches the mid-price when the spread is an even multiple of the tick size. Otherwise, it is adjusted to either half a tick above or below the mid-price, depending on the specific market conditions. For more details on this mechanism and its implications on market dynamics, see subsequent paragraph.

opposite side to fill the new price. Otherwise, the reference price stays at its value, and the levels don't change. The parameter θ is calibrated in order to reproduce the mean reversion ratio, as discussed in (Huang et al., 2015).
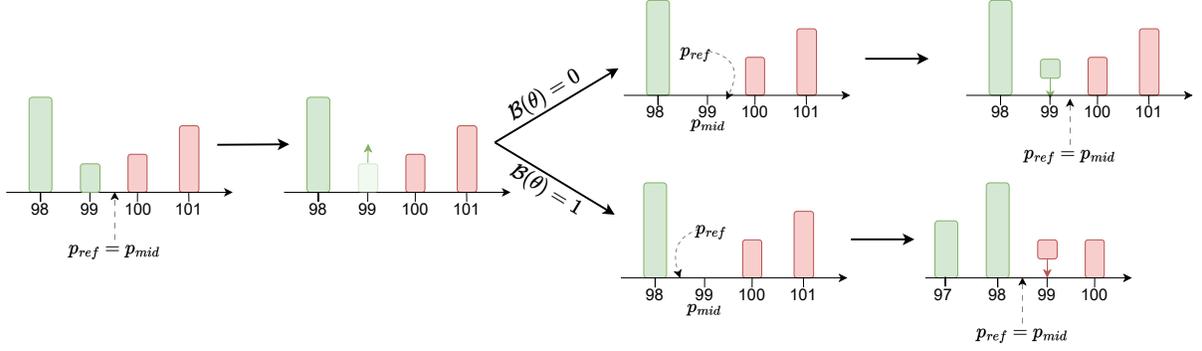

Figure 2: Example of price update scenario.

For a set $\mathcal{E}$ of events at a given queue i, $\mathcal{E} = (e_k = (\eta_k, q_k, \Delta t_k))_k$, the likelihood of the QR model can be expressed as:

$$\mathcal{L}(\lambda | \mathcal{E}) = \prod_k e^{-\lambda^{\text{global}}(q_k) \Delta t_k} \lambda^{\eta_k}(q_k) \tag{1}$$

where:

- η_k is the type of the $k - th$ event, $\eta \in \mathcal{T}, \mathcal{T} = \{L, C, M\}$ is the set of possible event types.

- q_k is the size of the queue just before the $k - th$ event.

- Δt_k is the time difference between events $k - 1$ and k.

- $\lambda^{\text{global}}(q_k) = \sum_{\eta \in \mathcal{T}} \lambda^{\eta}(q_k)$, the intensity of arrival of an order when the size of the queue is q_k.

4

For a given size of the queue n measured in Average Event Size[2] (AES), the model admits closed-form solutions:

$$\hat{\Lambda}(n) = \big(\text{mean}(\Delta t_k \mid q_k = n)\big)^{-1}$$

$$\hat{\lambda}^L(n) = \hat{\Lambda}(n)\frac{\#\{e_k \in \mathcal{E} \mid \eta_k = L, q_k = n\}}{\#\{e_k \in \mathcal{E} \mid q_k = n\}}$$

$$\hat{\lambda}^C(n) = \hat{\Lambda}(n)\frac{\#\{e_k \in \mathcal{E} \mid \eta_k = C, q_k = n\}}{\#\{e_k \in \mathcal{E} \mid q_k = n\}}$$

$$\hat{\lambda}^M(n) = \hat{\Lambda}(n)\frac{\#\{e_k \in \mathcal{E} \mid \eta_k = M, q_k = n\}}{\#\{e_k \in \mathcal{E} \mid q_k = n\}}$$

In our study, we base our experimental results on Bund futures data. Bund futures are well-known for their high liquidity and the abundance of data they provide, making them ideal for detailed analysis. The Bund is a large tick asset, for which the spread is usually of unitary tick, which helps reduce the impact of large spreads on model behavior. This is quite different from the assets used in Huang et al. (2015), allowing us to test the model in a market setting with different characteristics.
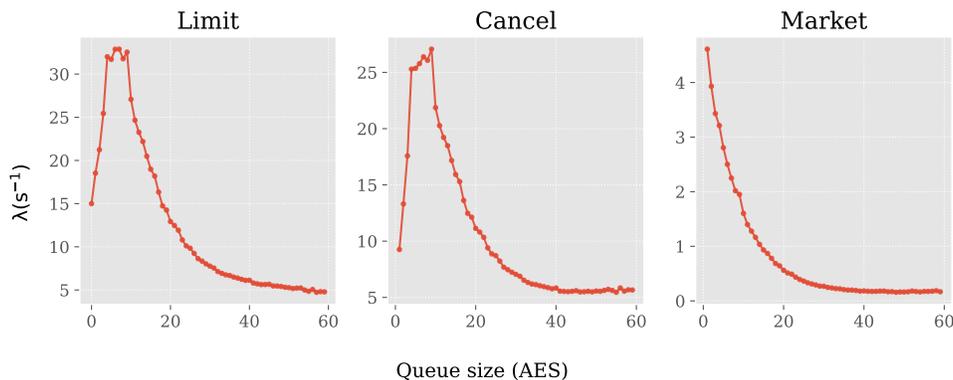

Figure 3: Results of queue-reactive model on Bund futures - Intensities as functions of queue size (best levels).

Figure 3 shows the calibrated intensities on our data, one can conclude that:

- Intensities vary significantly with the size of the queue, which justifies the choice of this model over a simple Poisson model.

- The limit intensity generally decreases, except when queue sizes are *tiny*. This phenomenon can be understood by considering the behavior of Market Participants

---

[2]This is the average of all events sizes occurring regardless of their type. Huang et al. (2015) propose also the usage of Average Trade Size (ATS) as a normalizing factor, which accounts only for the size of trade events.

(MPs). MPs may perceive it as risky to enter a level with an extremely low volume, up until a certain threshold (typically between 7 AES to 10 AES) is reached. At this threshold, there is a noticeable spike in limit orders, which can be attributed to the MPs' pursuit of priority. However, as the size of the queue grows, this pursuit diminishes.

- The same behavior is observed for cancel orders. However, it has the opposite interpretation; this may be because MPs who do not want to be executed retire when the size of the queue is around 10 AES. After that, market participants generally prefer to be executed rather than cancel and move to a deeper level. This contradiction in behavior with the limit intensity may indicate that there are two major types of market participants: one type always seeks to be executed as quickly as possible without paying the spread (without being aggressive), and others who prefer to seek priority without being executed, possibly waiting for better opportunities. This hypothesis of a fragmented market has also been adopted by other research works such as De Grauwe et al. (1995) or Lux and Marchesi (1998).

- The trading rate decays with the available volume at the best level. This phenomenon is easily explained by market participants *rushing for liquidity* when it is scarce, and *waiting for a better price* when liquidity is abundant, similar results are noticed on the equity markets in the Huang et al. (2015) study.

Now that we have the intensity of arrival of orders, and their types, the last and only missing information about the order is its size. We explore two methods of orders sizing:

- Each order is assigned a uniform size, determined by rounding up the average event size ($\lceil AES_i \rceil$) for each level i. This approach, referred to as **QRU** (QR model with Unitary order sizes), aligns with the model used in (Huang et al., 2015).

- The size of each order is sampled from the stationary distribution of order sizes for its respective level, denoted as **QR**.

2.2. Queue-Reactive Model Extensions

In addition to these two variants of the QR model, we propose two other extensions heavily inspired by it. The goal is to examine the impact of incorporating a more sophisticated model for order sizes on the simulator performances. The first extension involves a modified QR model that includes five possible event types, by adding two additional types, *cancel_all* and *market_all*, that model the complete consumption of the queue. This aims to simulate instances where the size of the consumption order precisely matches the queue size. The other extension involves incorporating the order size as a variable in determining the nature of order arrival, similar to its influence by event type.

2.2.1. Five-Type Queue-Reactive model (**FTQR**)

As mentioned earlier, and as will be discussed in the results section later, one of the main limitations of the queue-reactive model is its inability to reproduce historical market volatility. This is primarily due to a slower progression of the LOB when the size of one of the best queues is small, compared to historical progression. In fact, as the sizes of the orders are drawn from the stationary distribution of sizes, which is very skewed towards small sizes, it leads to a very slow decrease of the queue consumption, hence, fewer price changes and low volatility.

In fact, Figure 4 presents the stationary distribution of order sizes truncated at 30 lots ($\approx 5AES$)[3] (*stationary*), and the conditional distribution of order sizes when the size of the queue is 30 lots (*conditional*). It is an example of the state of order sizes when the queue size is relatively low. This illustration highlights the disparity between the two distributions, with the conditional distribution leaning towards larger values and, correspondingly, greater queue consumption.

In particular, there is a pronounced surge of orders that deplete 100% of the queue, predominantly for market orders. The frequency of such orders is more than 40 times higher than that in the stationary queue size consumption ratio. This significant discrepancy provides an explanation as to why queue consumption is markedly slower when stationary sizes are utilized. Especially after showing the link between queue consumption and price dynamics in Figure 2. This, in turn, elucidates why the simulated price exhibits considerably less volatility compared to actual prices.
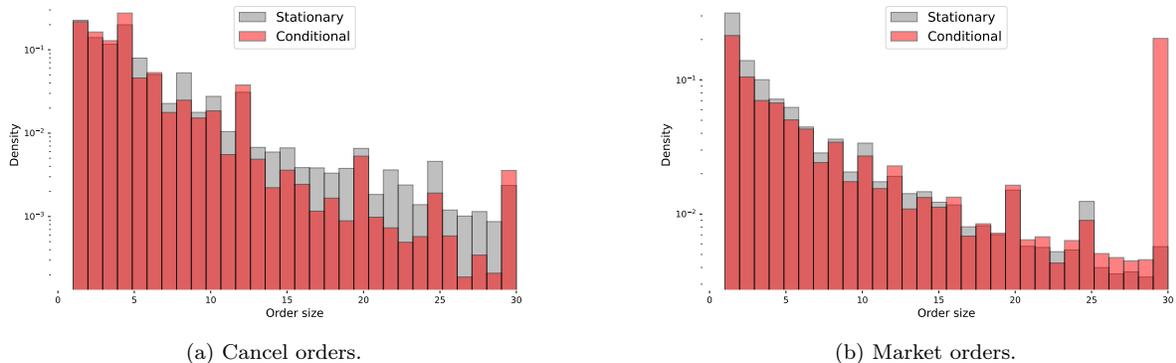

| (a) Cancel orders. | (b) Market orders. |

Figure 4: Comparison of stationary distribution of order sizes up to 30 (in gray) with conditional distribution when queue size equals 30 (in red), presented on a logarithmic scale.

These insights motivate a first extension of the QR model that take into account the phenomenon of complete queue consumption when the size of the queue is small. In such situations, two specific events tend to occur:

- Market participants seize the available quantity and aggressively execute it (using

---

[3]In the context of BUND Futures, a queue comprises multiple contracts, each valued at 100,000 Euros. The distribution of order sizes truncated at 30 lots corresponds to a financial value of approximately 3 million Euros, which is about five times the Average Event Size (AES).

market orders).

- Market participants who possess these scarce quantities feel uncomfortable being alone in the queue and cancel all their orders. Another way of seeing it is that when market participants are alone in the queue, it is *free* to cancel their order without worrying about losing priority in the queue.

In this context, we propose adding two additional events that correspond to the complete consumption of available quantity, either through trades or cancellations. The likelihood of this new model is similar to that presented in Equation 1 by including *market_ all* and *cancel_ all* events in $\mathcal{T}$, the set of possible event types.
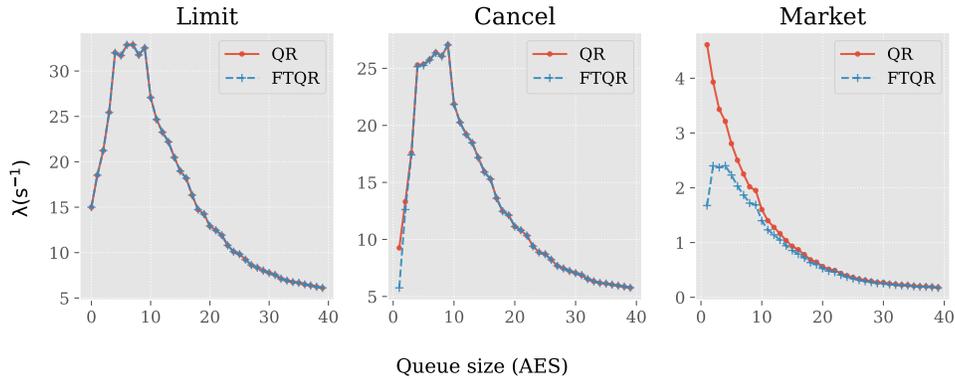

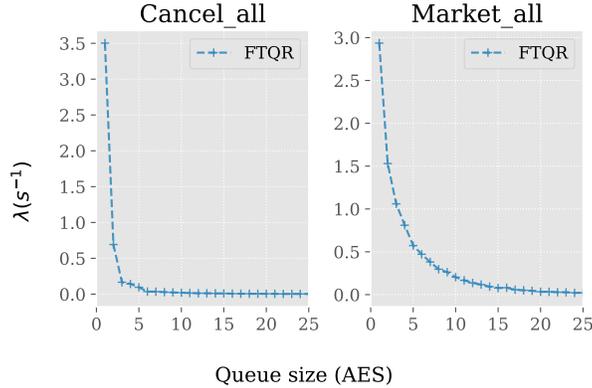Figure 5: Comparison of QR and FTQR intensities on Bund future data.

Figure 6: Calibrated intensity functions for additional types of the FTQR model applied to Bund future data, truncated at values less than 25 lots. Intensities for larger queue sizes are negligible.

Figures 5 and 6 above show the obtained intensities of the five types queue-reactive model and compare them with the vanilla QR model intensities, revealing the following observations:

- **Limit orders:** The limit order intensities are identical in both models, as the adjustment does not affect limit orders.

- **Cancel orders:** The cancel order intensities are also the same, except for smaller queue sizes where the intensities of the adjusted model are lower. This difference originates from the additional cancel type that consumes the entire level. Nevertheless, the impact of such cancel orders on overall liquidity is negligible compared to other cancel orders.

- **Market orders:** This order type is the most influenced by the adjustment. We notice that the new intensities are lower than the original ones, particularly for smaller queue sizes, as these situations provide more opportunities for rarity chasing.

- **Market_all orders:** The new intensities compensate for the earlier ones, particularly for small queue values, reflecting a significant value. This finding indicates that many orders consume liquidity at a high rate when this liquidity is scarce. The intensity of market_all orders decays with the queue size.

- **Cancel_all orders:** This order type exhibits a behavior similar to the previous one. It is mainly interpreted as the intensity of arrival of events where a market participant, who is alone in the queue, completely withdraws from the market.

*2.2.2. Size-Aware Queue-Reactive model (**SAQR**)*

In this section, we aim to explore the impact of order sizes on the simulation of limit order books. In fact, research interest in the field of limit order book simulation has been particularly focused on determining the rate of arrival of events and their types. However, the size of orders has received less attention.

In this section, we propose an adaptation of the QR model that takes into account the size of orders. We achieve this by changing the possible values of $\mathcal{T}$ in Equation 1 to a 2-dimensional order type, $\mathcal{T} = \{L, C, T\} \times \mathbb{N}^*$. This means that the intensity nature not only specifies the type of the order but also its size. This adjustment does not change the formulation of the likelihood. And the solution, for (η, s) and a queue size n, is:

$$\hat{\lambda}^{(\eta,s)}(n) = \hat{\Lambda}(n) \frac{\#\{e_k \in \mathcal{E} \mid \eta_k = \eta, s_k = s, q_k = n\}}{\#\{e_k \in \mathcal{E} \mid q_k = n\}}$$

where s_k is an additional model variable and represents the size of the $k-th$ order.

Figure 7 presents the intensity matrix for the three types of orders: limit, cancel, and market orders (averaged for each queue size). It is evident from the figure that while the distribution of limit order sizes remains relatively consistent with respect to the queue size, the same cannot be said for cancel and market orders.

In particular, the diagonal values in the intensity matrix, which represent orders that consume the entire available liquidity, are noticeably dominant. This dominance is especially pronounced for market orders, confirming the tendency of these orders to consume the entire quantity available at the best levels.
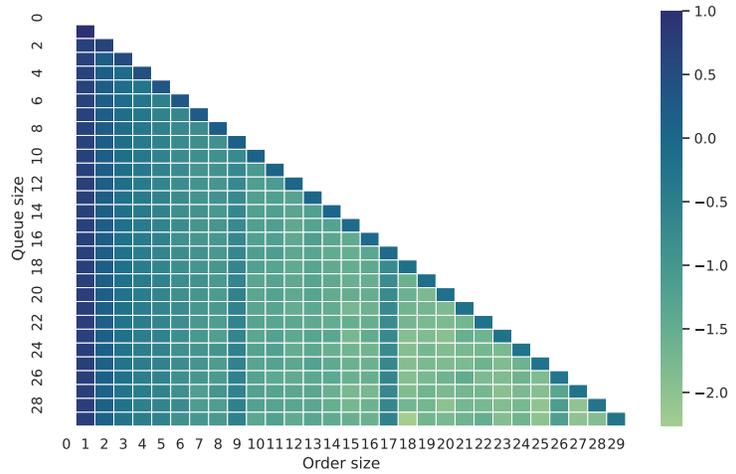
Interestingly, this phenomenon is more prevalent when the available quantity is relatively low. This observation highlights the fact that the distribution of order sizes is not uniform, but rather, it varies significantly depending on the size of the queue. This
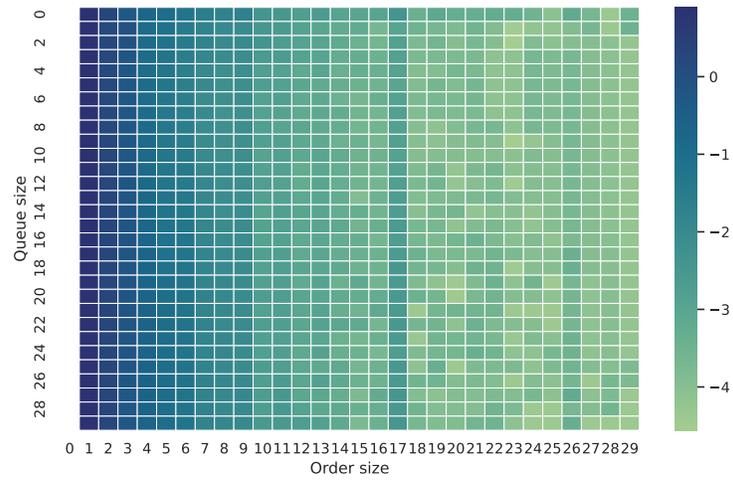
(a) Limit orders
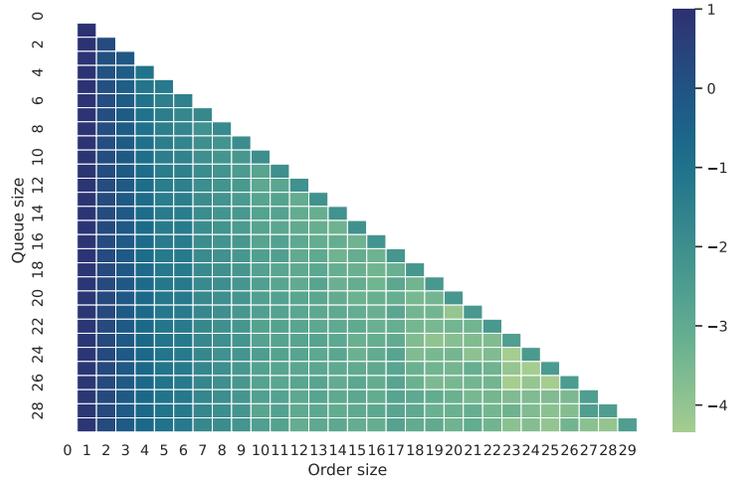
(b) Cancel orders

(c) Market orders

Figure 7: Matrix heatmap of normalized calibrated intensities On Bund future data (logarithmic scale). Normalization is done such that intensities sum to one for each row.

nuanced understanding of order size distribution is crucial for accurately simulating and analyzing the dynamics of the limit order book.

Furthermore, one can notice in the conditional market intensities some particularly *salient* intensities at 9 AES, 17 AES and 26 AES. These are buckets including orders of sizes 50 lots, 100 lots and 150 lots, respectively. These round numbers are generally more occurrent in historical data than other order sizes, hence the high intensity of arrival of orders of these sizes.

2.3. Hawkes Models

Hawkes processes, named after Alan Hawkes (Hawkes, 1971), are a class of self-exciting and mutually-exciting point processes. Unlike Poisson processes, they dispense with the independence-of-increments property, allowing past events to influence the likelihood of future events. This characteristic makes them particularly adapted for modeling clustering and contagion phenomena, which are prevalent in social media networks, seismology, and financial markets.

Mathematically, a Hawkes process is defined by its intensity function, $\lambda(t)$, which depends on the entire history of the process. The intensity function is often formulated as:

$$\lambda(t) = \mu + \int_{-\infty}^{t} h(t-s)dN(s)$$

where μ is the baseline intensity, $h(t-s)$ is a memory kernel describing the influence of past events, and $N(s)$ is the count of events up to time s.

The complexity of financial markets, with multiple interacting variables, necessitates a more comprehensive approach to capture these dynamics accurately. This leads us to the multivariate Hawkes process, which extends the basic concept to encompass multiple types of interdependent events.

Mathematically, a multivariate Hawkes process is defined by a set of intensity functions, $\lambda_i(t)$, each depending on the entire history of the process. The intensity function for the i-th component of the process at time t can be expressed as:

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{n} \int_{-\infty}^{t} h_{ij}(t-s)dN_j(s)$$

where μ_i is the baseline intensity for the i-th component, $h_{ij}(t-s)$ is a memory kernel describing the influence of past events of type j on the current rate of type i, and $N_j(s)$ is the count of type j events up to time s. This model captures not only the self-excitations ($j = i$) of events, but also the cross-excitation ($j \neq i$) between events.

In the context of limit order book (LOB) modeling, Hawkes processes have been instrumental in describing the flow of orders. These models view different order types as distinct events in a multivariate point process, capturing the dynamics of market reactions to these orders and their interrelations. Significant developments in LOB modeling include contributions from Zhao (2010) and Toke (2011), who have expanded upon the Cont et al. (2010) framework. Zhao's empirical analysis of crude oil futures led to the

rejection of the assumption that market event inter-arrival times were independent and identically distributed, challenging the conventional use of independent Poisson processes. Instead, Zhao (2010) introduced a Hawkes process, as suggested by Bauwens and Hautsch (2009), to model the arrival rate of market events as a function of recent order arrival rates and quantities. This approach better represented periods of clustered high and low arrival rates, aligning more closely with empirical observations (Ellul et al., 2003; Hall and Hautsch, 2006). Similarly, Toke (2011) replaced the Poisson processes in the Cont et al. (2010) model with multivariate mutually-exciting Hawkes processes, tailoring them for each type of market event.

However, there remains a need for models that also account for the state of the LOB, incorporating variables such as prices and bid-ask spreads. This has led to the development of more complex models that blend the strengths of Hawkes processes with other modeling approaches, offering a more comprehensive view of the dynamics within financial markets.

In our study, we follow a methodology akin to that of Zhao (2010) and Toke (2011), aiming to compare it to the various versions of the queue-reactive model. To this end, we calibrate a Hawkes process to develop our simulation framework. Our parametrization strategy is detailed as follows:

- We treat all events occurring at the best limits as a single, six-dimensional multivariate process. This approach accounts for three events at each best price, allowing us to capture not only the excitation between events within the same queue but also the interdependencies between the bid and ask sides of the market.

- Exponential kernels are employed to model the time-dependent influence of past events on future occurrences. Which means, the memory kernel functions are of the form $h_{ij}(t) = \alpha_{ij} e^{\beta_{ij} t}$.

- We adhere to the same price change paradigm as established in the queue-reactive model, ensuring consistency and comparability in our analyses.

For the calibration of this model, we utilized the *tick* package of Bacry et al. (2017). Once calibrated, the same package facilitated the simulation of event sequences and inter-arrival times. These simulations were integral in constructing the order flow, which was subsequently incorporated into our matching engine to generate the final simulated order book.

In parallel with the queue-reactive model, we examine two variants of the Hawkes-based model in relation to order sizes. The first variant, referred to as **Hawkes_U**, operates on the assumption of unitary order sizes. The second variant, dubbed **Hawkes_S**, utilizes the stationary distribution of order sizes. This bifurcation allows for a nuanced examination of the models' performance under different assumptions about market order size models.

3. Results

3.1. Data & Calibration Paradigm

The data used for calibration is from Euro-Bund Futures (FGBL)[4] covering active trading days on the year 2021. Data is filtered into periods from 9 AM to 6 PM. The maximum depth considered is five levels on each side.

Level	#L ($\times 10^6$)	#C ($\times 10^6$)	#M ($\times 10^4$)	AES	AIT (ms)
1	119.6	115.6	756	24.92	56
2	24.09	23.84	2.88	5.15	272
3	15.76	15.29	0	4.87	370
4	10.72	10.45	0	5.98	468
5	11.12	10.76	0	4.11	523

Table 1: Descriptive statistics about events for each level.

Table 1 provides statistics about the data utilized for model calibration. It displays the number of observations for each event type, the Average Event Size (AES), and the Average Inter-event Time (AIT) for each queue. Notably, the table underscores the high frequency of events impacting queues close to the mid-price, especially the best prices.

Additionally, our dataset includes updates to the limit order book and the daily trade history. However, the models necessitate training with order flow, representing the sequence of events in the limit order book. To obtain this order flow, we carefully constructed it by discerning whether each update indicates liquidity provision (via a limit order) or liquidity consumption (through cancel and market orders). Further details on order flow construction are available in Bodor and Carlier (2023). Additionally, a matching engine was developed to pair orders produced by each model, subsequently reflecting them in a simulated order book.

This data is further pre-processed following the methodology in Huang et al. (2015). Specifically:

- Queue sizes are quantized based on the average event size. If the size of queue i is denoted as q_i and its average event size is AES_i, then the updated queue size is given by $q_i \leftarrow \lceil \frac{q_i}{AES_i} \rceil$.

- The data is divided into periods with a constant reference price. The interarrival time counter for events is reset to 0 for all queues when there's a change in the reference price.

---

[4]Euro-Bund Futures are derivative contracts based on 10-year German government bonds (Bund) and traded on the Eurex Exchange. Each futures contract, with a notional coupon rate of 6%, has a remaining term to maturity of 8.5 to 10.5 years on the delivery day.

- The parameter that controls the reference price movements is $\theta = 0.7$. This probability has been calibrated to reproduce the observed ratio of price continuations to price alternations.

For model intensity calibration and subsequent LOB simulations, we:

- Utilize the calibrated intensities to determine both the time of arrival and the type of new orders.

- For modeling price changes, we adhere to the aforementioned paradigm. However, we refrain from redrawing the LOB from its stationary distribution to exclusively replicate the endogenous market behavior.

- Upon a change in the reference price, the size of the queue for newly appearing prices is sampled from their respective stationary distributions.

3.2. Stylized Facts

In our quest to assess the efficiency of different market simulators, we advocate for their comparison against a set of established stylized facts from the market. These facts span micro to macro-level empirical findings, particularly tailored for the Bund futures market. As detailed in Bodor and Carlier (2023), we delve into an intricate analysis of stylized facts pertinent to German bond futures, spotlighting Schatz, Bobl, Bund, and Buxl, grounded on exhaustive limit order book data. This exploration discerns both commonalities and unique traits across these futures, shedding light on market dynamics and pointing towards the refinement of market simulation tools. Our primary focus is to benchmark our models against the stylized facts observed in the Bund futures market (Bodor and Carlier, 2023), as detailed in the following sections.
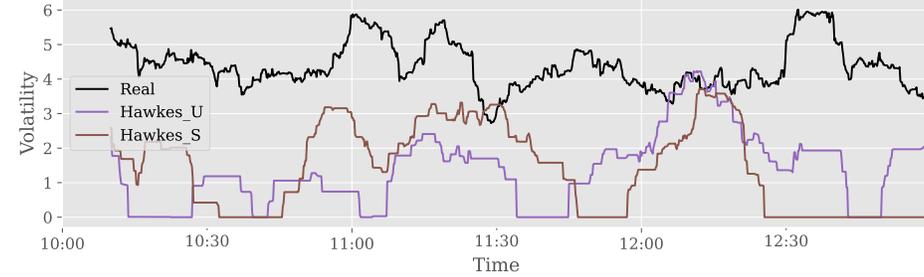
3.2.1. Price dynamics and volatility

A fundamental feature of an effective simulator is its ability to simulate realistic price dynamics and replicate the volatility inherent to the target market. Figure 8 illustrates the price dynamics of both historical data and simulations generated using the different models.

It becomes apparent that the queue-reactive model, particularly the unitary sized variant, exhibits slower dynamics than the actual market, coupled with reduced volatility. Conversely, the FTQR model seems to engender a markedly noisier market environment. Meanwhile, the SAQR model's performance closely aligns with real price patterns. Regarding the Hawkes-based simulation models, they display characteristics similar to those of the standard queue-reactive model. Notably, these simulators also tend to mirror the slower market dynamics observed in the queue-reactive model, contrasting with the quicker pace of real-world markets.
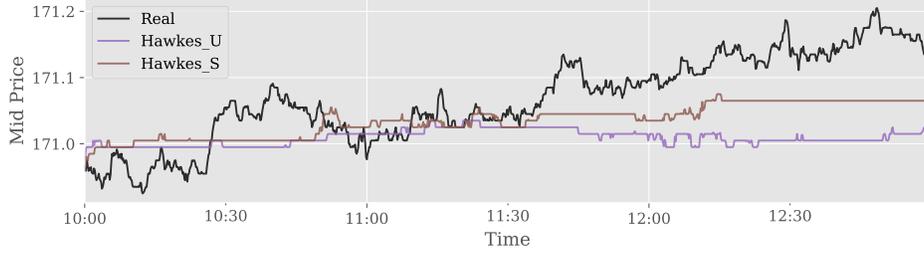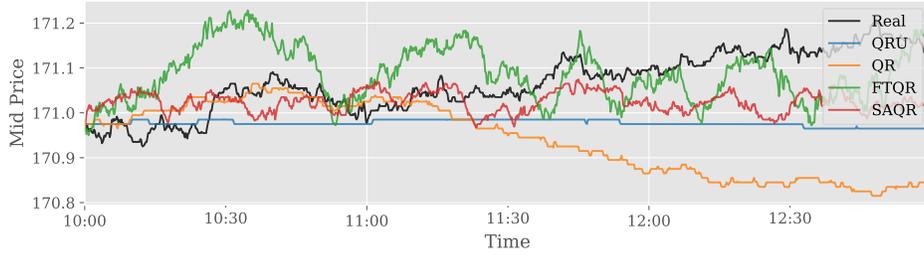
(a) Mid-price

(b) Annualized volatility over windows of size 10 minutes

Figure 8: Mid-price dynamics and annualized volatility plots (1 second sampling frequency).

More generally, Table 2 presents a comprehensive comparison of price volatility across different days of data and distinct periods of the day. This comparison employs two metrics:

- Relative difference: $\text{mean}\left(100\% \times \frac{\sigma_s - \sigma_r}{\sigma_r}\right)$

- Quadratic error: $\text{mean}\left((\sigma_s - \sigma_r)^2\right)$

Here, σ_s and σ_r denote the annualized volatility of the simulated and real markets, respectively.

The subsequent table delineates the comparative volatility of simulated markets vis-à-vis the real market. Notably, the SAQR model persistently outperforms other models across all periods of the day, reflecting the smallest average relative difference and quadratic error. This sharply contrasts the QRU model, which exhibits considerably diminished volatility relative to the real market.

The performance trajectories of the QR and FTQR models warrant special attention, as they fluctuate based on the time of day. The QR model exhibits closer adherence to real market dynamics during the 10 AM to 2 PM bracket, generally perceived as the "calm" periods of the day. Conversely, the FTQR model manifests superior performance from 3 PM to 6 PM, periods known for high market activity driven by events like closing auctions and options expiry.

	9-18		10-14		15-18	
	Relative Difference (%)	Quadratic Error	Relative Difference (%)	Quadratic Error	Relative Difference (%)	Quadratic Error
QRU	-91	18.37	-90	14.97	-91	21.50
QR	-52	7.34	-48	5.17	-54	9.56
FTQR	69	8.88	82	10.71	49	8.12
SAQR	**1**	**1.35**	**8**	**0.82**	**-3**	**2.33**
Hawkes_U	-72	12.46	-70	9.64	-73	15.16
Hawkes_S	-69	11.70	-67	8.96	-71	14.33

Table 2: Statistics about volatility of simulated markets compared with real market on different periods of the day.

3.2.2. Transactions magnitude and market activity

An essential aspect of limit order book simulators is their ability to accurately capture the magnitude of transactions and orders. This entails the capability of the simulator to replicate a similar order of magnitude of traded volumes within a given time window. As illustrated in Figure 9, which shows that the SAQR model exhibits the highest proficiency in replicating traded volumes during typical trading periods of the day. Both the QR and QRU models demonstrate a relatively low transactions volume. However, the FTQR model exhibits a significantly higher rate. The Hawkes-based models are also quite accurate at reproducing a relatively similar market activity, showcasing the importance of excitation between events for the verification of this stylized fact.
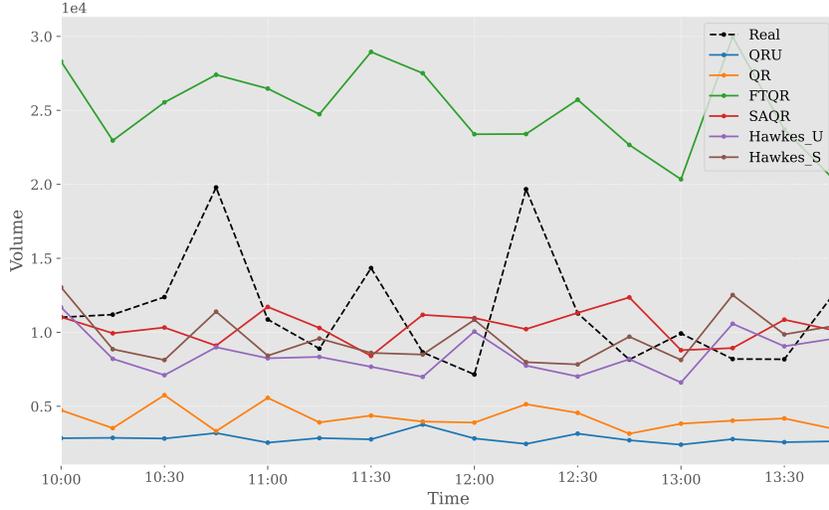
Figure 9: Traded volumes on windows of size 10 minutes.

	9-18		10-14		15-18	
	Relative Difference (%)	Quadratic Error ($\times 10^9$)	Relative Difference (%)	Quadratic Error ($\times 10^9$)	Relative Difference (%)	Quadratic Error ($\times 10^9$)
QRU	-80	2.02	-76	1.32	-82	2.81
QR	-70	1.55	-65	0.95	-74	2.26
FTQR	49	0.85	76	1.38	**30**	**0.63**
SAQR	**-26**	**0.36**	**-13**	**0.14**	-36	**0.62**
Hawkes_U	-40	0.60	-28	0.33	-48	0.97
Hawkes_S	-32	0.41	-19	0.21	-41	0.80

Table 3: Statistics about traded volumes of simulated during different periods of the day.

Table 3 provides a comprehensive comparison between the traded volumes of the simulated market and the real market. Consistently, the SAQR model emerges as the closest to real data, exhibiting the smallest quadratic error across all day periods, followed by the Hawkes-based model, that show also quite good results, with a notable improvement of the Hawkes_S model over Hawkes_U. Interestingly, the QR model outperforms the FTQR model during the periods from 10 AM to 2 PM, while the FTQR model appears to be better suited for replicating the high market activity observed from 3 PM to 6 PM.

3.2.3. Order book queue size distribution

An additional noteworthy empirical observation pertains to the distribution of available liquidity in the market. Numerous studies have not only acknowledged the consistency of this distribution but also reported the successful fit of the Gamma distribution to it (Abergel et al., 2016; Vyetrenko et al., 2020; Bodor and Carlier, 2023).

In Figure 10, the empirical distribution of queue sizes at the most favorable levels is effectively captured. All models show fit accurately the Gamma distribution, highlighting

17

their capability in modeling queue dynamics. Notably, the SAQR and FTQR models demonstrate a particularly strong alignment with actual market data, underscoring their accuracy in reflecting real market conditions. In contrast, the Hawkes-based models do not mimic the queue size distribution as accurately. They tend to simulate many limit orders, regardless of the queue's actual size. This results in an unrealistic pile-up of orders, which does not match what we usually see in real markets. These observations highlight the strengths of the queue-reactive model. It adjusts the flow of orders – both incoming and outgoing – based on the size of the queue. This method leads to a more accurate representation of market behavior, especially in terms of queue size distribution.
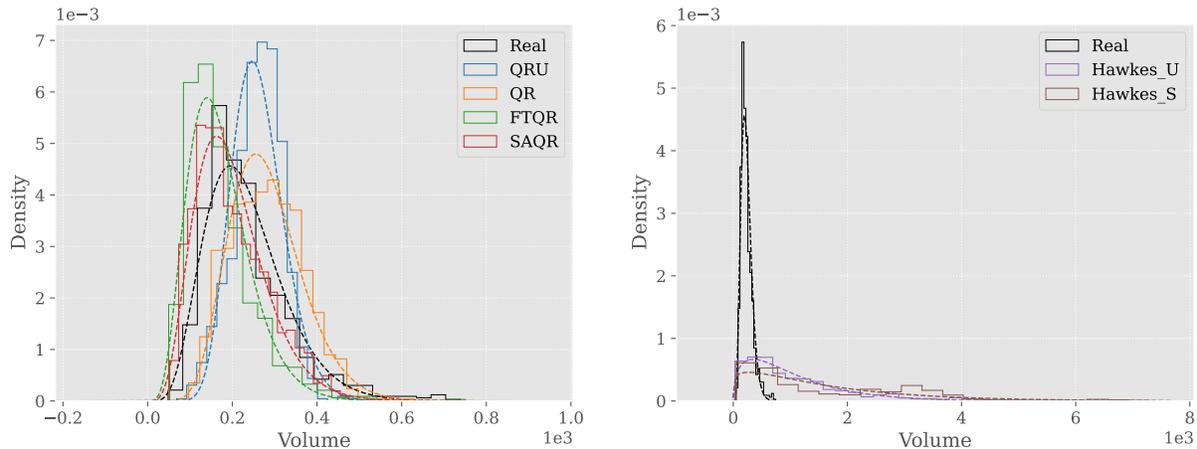

Figure 10: Distribution of order book volumes and Gamma law fit.

In order to compare these distributions across multiple days of data, we adopt the Kolmogorov-Smirnov statistic as a distance between distributions. The Kolmogorov-Smirnov (KS) test is a nonparametric test used to determine the similarity between two probability distributions. The test statistic of the KS test is defined as:

$$D_n = \sup_x |F_n(x) - F(x)|$$

where $F_n(x)$ is the empirical cumulative distribution function (CDF) of the observed data and $F(x)$ is the CDF of the reference distribution or theoretical model. The value of D_n lies between 0 and 1. The smaller the value of D_n, the more similar the observed data distribution is to the reference distribution. Therefore, the KS statistic can be used to measure the similarity between two distributions.

Table 4 presents the descriptive statistics of the Kolmogorov-Smirnov statistic, comparing the real and simulated queue size distributions from different days of calibration data. The table indicates that both the SAQR and QR models perform well on average, despite the fact that the volume distributions generated by these two models are quite distinct. This discrepancy can be attributed to the variability in market behavior, which aligns more closely with the QR distribution on some days and the SAQR distribution on others. Regardless, both models can be considered to validate the stylized fact.

18

In contrast, Hawkes-based models exhibit very wide distribution distances from the real distribution, further confirming that these models suffer from the queue size divergence problem, where, intrinsically, there is no guarantee of a return to equilibrium, unlike the queue reactive model.

	QRU	QR	FTQR	SAQR	Hawkes_U	Hawkes_S
mean	0.32	0.22	0.28	**0.20**	0.63	0.69
std	0.09	0.11	0.08	0.06	0.05	0.04
min	0.23	0.14	0.05	0.08	0.57	0.64
25%	0.27	0.15	0.26	0.15	0.60	0.66
50%	0.30	0.18	0.30	0.21	0.62	0.68
75%	0.31	0.22	0.33	0.26	0.65	0.71
max	0.68	0.61	0.37	0.28	0.78	0.81

Table 4: Kolmogorov-Smirnov statistic of the distribution of queue sizes in simulated markets compared to the real market, description over multiple days of calibration data.

3.2.4. Signature plot

The signature plot of a price is defined as the variance of price increments at a given lag value h, divided by h. In other words, it is defined as $\sigma_h^2 = \frac{\mathbb{V}(P(t+h)-P(t))}{h}$, $P(t)$ being the mid-price of the asset at time t. It is interesting to observe the evolution of this metric with the value of h, as this highlights the transition of price movements at different time scales. Figure 11 shows this transition for an example of one real day and simulated markets. While all the markets seem to exhibit the same shape of decrease in volatility as the time lag value increases, the SAQR model is the best at capturing the real signature plot at different time scales.
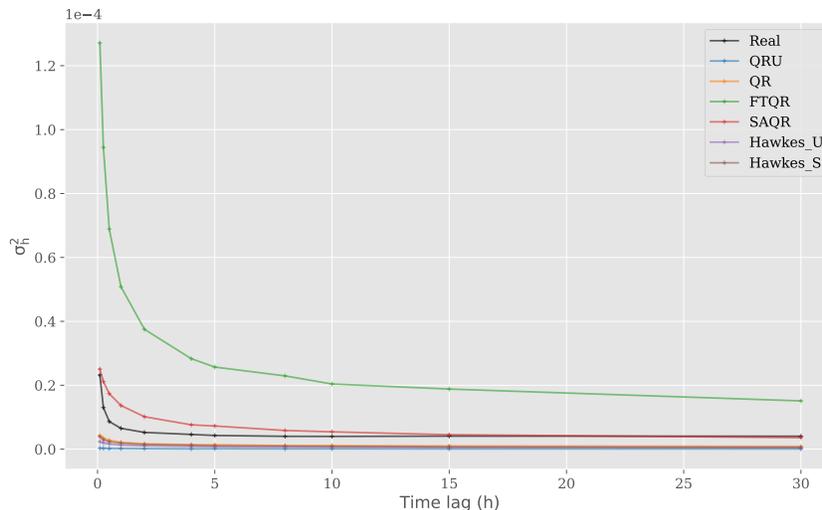

Figure 11: Signature plot.

More generally, the following table provides more statistics about comparing the signature plot of real data from different days to the signature plot of simulated data, the

distance between curves is defined as the sum of difference of values for each lag h. The table shows that the QRU, QR, Hawkes_U and Hawkes_S models generally underestimate the volatility while the FTQR heavily overestimates it. The SAQR is the model with the least quadratic error and least bias, confirming its ability to capture the real volatility at different scales. However, it is important to note that while the SAQR model excels in terms of quadratic error and bias, it still struggles to perfectly capture the exact decaying profile of the real data, especially at smaller time scales.

	Relative Difference (%)	L^2 Norm $\times 10^{-5}$
QRU	-9.80	2.89
QR	-7.55	2.28
FTQR	41.32	9.99
SAQR	**4.43**	**1.05**
Hawkes_U	-8.57	2.09
Hawkes_S	-7.91	1.95

Table 5: Statistics over multiple days of calibration data - signature plot values.

3.2.5. *Distribution of returns*

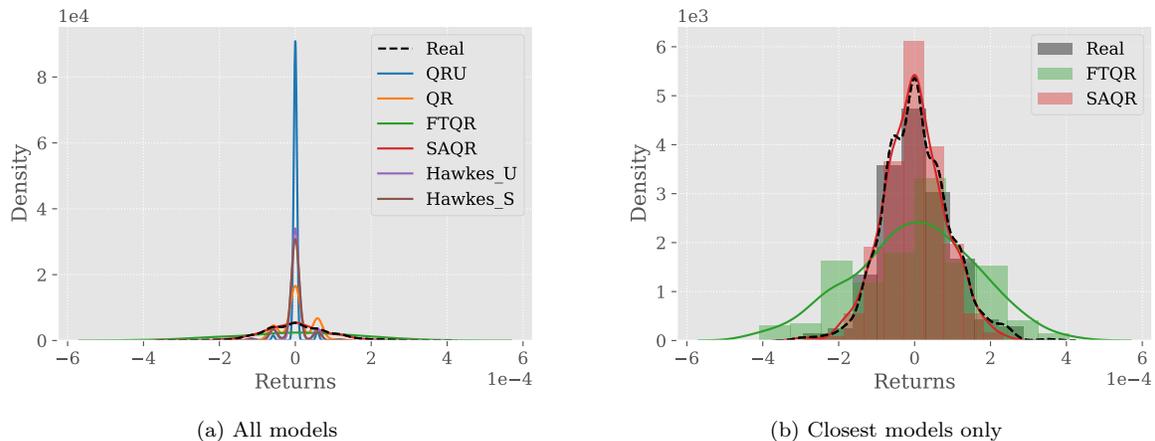(a) All models

(b) Closest models only

Figure 12: Distribution of logarithmic returns of prices $\log(\frac{P(t+\tau)}{P(t)})$ of real and simulated data.

Figure 12 displays a comparison of one daily return distributions between various simulated markets and the actual market. Part (a) of the figure compares the distributions of all models against the real market, while part (b) focuses on the best-performing models for enhanced clarity. This comparison reveals that the SAQR model's distribution of returns closely resembles that of the real market. In part (a), it's noticeable that the QR, QRU, Hawkes_U, and Hawkes_S models generate returns that are too concentrated around zero, reflecting a slower volatility than what is seen in the actual market. On the other hand, part (b) illustrates how the SAQR model's return distribution closely mirrors

the real market. This finding is corroborated by Table 6, which indicates that the SAQR and FTQR models have return distributions most similar to those of the real market, with the SAQR model being the closest.

	QRU	QR	FTQR	SAQR	Hawkes_U	Hawkes_S
mean	0.52	0.35	0.21	**0.16**	0.43	0.42
std	0.04	0.05	0.05	0.03	0.03	0.03
min	0.43	0.25	0.12	0.11	0.39	0.37
25%	0.49	0.31	0.17	0.14	0.41	0.40
50%	0.51	0.35	0.20	0.15	0.43	0.42
75%	0.55	0.37	0.23	0.17	0.46	0.44
max	0.60	0.44	0.35	0.25	0.50	0.48

Table 6: Kolmogorov-Smirnov statistic of the distribution of returns in simulated markets compared to the real market. Description over multiple days of calibration data

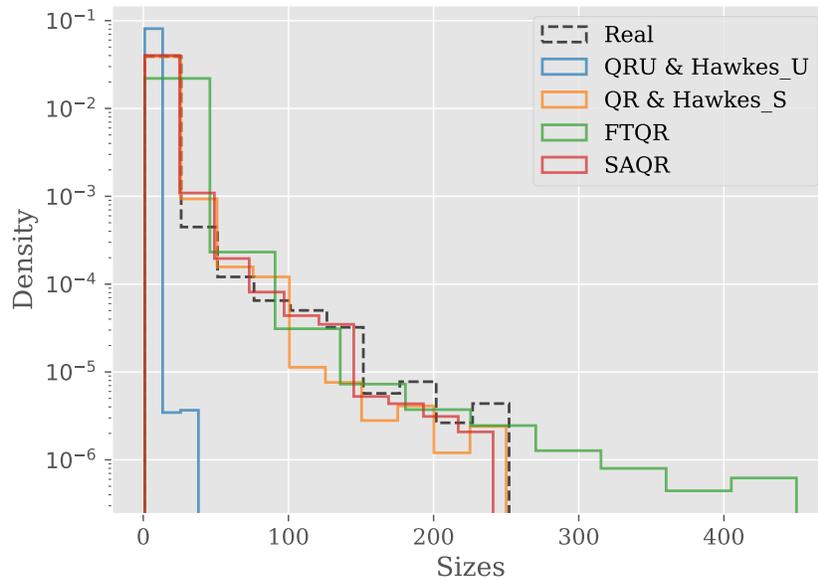### 3.2.6. Order sizes distribution

Figure 13: Distribution of order sizes (the y-axis is in logarithmic scale).

Figure 13 provides insights into the distribution of order sizes in both real and simulated markets. Several observations can be made:

- As anticipated, the QRU and Hawkes_U models perform poorly in reproducing this stylized fact. All orders in these models are unitary and clustered around the average event size (AES), resulting in a limited range of order sizes, and completely different distribution.

- The QR and Hawkes_S models exhibit a good fit to the historical distribution of order sizes. This outcome is expected since these models sample, by construction, order sizes from the stationary distribution of real market order sizes.

- Conversely, the SAQR model demonstrates a strong capability in reproducing the distribution of order sizes. It is important to emphasize that, in contrast to the QR model, this capability is not inherent to the SAQR model's design. The order sizes in the SAQR model are drawn based on a distribution conditional on the queue size. This feature allows the SAQR model to capture the empirical order size distribution more accurately, accounting for the realism of the queue sizes dynamics of this model.

- The FTQR model, while similar to the QR model in terms of order size picking paradigm, presents a distribution slightly shifted towards larger sizes. This shift is primarily due to the introduction of the two event types *cancel_all* and *trade_all* which encourage the placement of larger orders. This could also justify the increased volatility and activity observed in this model.
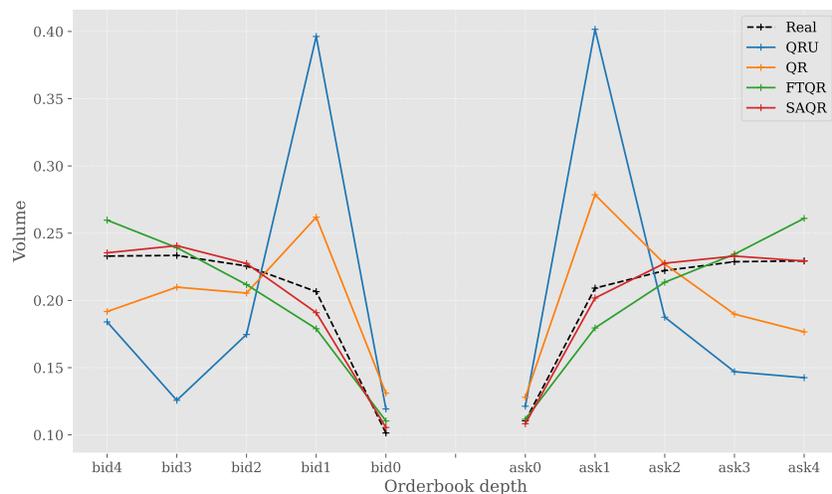
3.2.7. Order book shape

Figure 14: Order book mean volumes (averaged).

Figure 14 illustrates the average volume available at each queue up to the fifth level, normalized by the average of all queue sizes. This figure indicates that the limit order book is generally more populated at deeper levels, a phenomenon also observed on other markets (Abergel et al., 2016; Chakraborti et al., 2011), especially on the Bund Bodor and Carlier (2023).

This stylized fact is particularly evident in the FTQR and QR models. However, it appears less applicable to the QRU and QR models. This difference underscores the superior ability of the FTQR and SAQR models to accurately simulate queue size distributions at various price depths. The Hawkes-based models are absent in the figure as they simulate only best prices order flow.

22

3.2.8. Summary

Additional stylized facts are detailed in Appendix A. Tables 7 and 8 present a comprehensive list of stylized facts examined in our study, along with a performance summary for each model concerning these facts. We have used a color-coded system for clarity: green signifies that a model successfully captures a stylized fact, orange denotes partial accuracy, red indicates a miss, and black dash means irrelevant. The SAQR model emerges as a clear frontrunner, corroborating the most stylized facts and thereby underscoring its proficiency in mimicking real market dynamics. Conversely, the QRU and Hawkes_U models lag behind, emphasizing the significance of incorporating order sizes in market simulations. To sum up, accurately modeling order sizes is instrumental in enhancing the authenticity of simulated markets and bolstering the efficacy of the queue-reactive model.

Index	Stylized fact name
1	Distribution of order sizes
2	Power-law of order sizes distribution
3	Signature plot
4	Distribution of available volumes in the queue
5	Price dynamics and volatility
6	Long range dependency
7	Distribution of returns
8	Order book shape
9	Absence of autocorrelation
10	Traded volumes in a fixed window
11	Weibull fit of interarrival time of trades
12	Excitation between events

Table 7: Table of stylized facts.

Model	1	2	3	4	5	6	7	8	9	10	11	12
QRU	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	✓	✗
QR	✓	✓	✗	✓	✗	✓	✗	✗	✓	✗	✓	✗
FTQR	✗	✓	✗	✓	✗	✓	✗	✓	✓	✗	✓	✗
SAQR	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
Hawkes_U	✗	✗	✗	✗	✗	✓	✗	-	✓	✓	✓	✓
Hawkes_S	✓	✓	✗	✗	✗	✓	✗	-	✓	✓	✓	✓

Table 8: Summary of stylized facts verified by each model.

4. Conclusion and Perspectives

In this paper, we delve into the queue-reactive model's various adaptations to simulate limit order book dynamics accurately, using empirical data, especially from German bond futures. Our research uniquely incorporates and demonstrates the critical importance of order sizes, alongside order types and arrival rates, in realistically modeling limit order book flows. This methodological enhancement is validated by aligning closely with numerous market stylized facts, emphasizing the model's ability to accurately replicate real market behaviors. Moving forward, we aim to further refine the model by integrating additional dynamics such as event excitation and intra-day seasonality, among others, ensuring a balance between simplicity and comprehensive market representation.

Acknowledgements

References

Á. Cartea, S. Jaimungal, J. Penalva, Algorithmic and high-frequency trading (2015).

S. N. Cohen, L. Szpruch, A limit order book model for latency arbitrage, Mathematics and Financial Economics 6 (2012) 211–227.

T. H. Balch, M. Mahfouz, J. Lockhart, M. Hybinette, D. Byrd, How to evaluate trading strategies: Single agent market replay or multiple agent interactive simulation?, arXiv preprint arXiv:1906.12010 (2019).

D. Byrd, M. Hybinette, T. H. Balch, Abides: Towards high-fidelity market simulation for ai research, arXiv preprint arXiv:1904.12066 (2019).

S. Vyetrenko, D. Byrd, N. Petosa, M. Mahfouz, D. Dervovic, M. Veloso, T. Balch, Get real: Realism metrics for robust limit order book market simulations (2020) 1–8.

S. Amrouni, A. Moulin, J. Vann, S. Vyetrenko, T. Balch, M. Veloso, Abides-gym: gym environments for multi-agent discrete event simulation and application to financial markets (2021) 1–9.

J. Lussange, I. Lazarevich, S. Bourgeois-Gironde, S. Palminteri, B. Gutkin, Modelling stock markets by multi-agent reinforcement learning, Computational Economics 57 (2021) 113–147.

J. Sirignano, R. Cont, Universal features of price formation in financial markets: perspectives from deep learning, Quantitative Finance 19 (2019) 1449–1459.

P. Zhang, H. Chen, Z. Liu, Z. Liu, Y. Yan, Integrating the data augmentation scheme with various classifiers for acoustic scene modeling, arXiv preprint arXiv:1907.06639 (2019).

A. Coletta, J. Jerome, R. Savani, S. Vyetrenko, Conditional generators for limit order book environments: Explainability, challenges, and robustness, arXiv preprint arXiv:2306.12806 (2023).

W. Huang, C.-A. Lehalle, M. Rosenbaum, Simulating and analyzing order book data: The queue-reactive model, Journal of the American Statistical Association 110 (2015) 107–122.

T. Mariotti, F. Lillo, G. Toscano, From zero-intelligence to queue-reactive: limit-order-book modeling for high-frequency volatility estimation and optimal execution, Quantitative Finance 23 (2023) 367–388.

P. De Grauwe, H. Dewachter, M. Embrechts, Exchange rate theory: chaotic models of foreign exchange markets (1995).

T. Lux, M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, Nature 397 (1998) 498–500.

A. G. Hawkes, Spectra of some self-exciting and mutually exciting point processes, Biometrika 58 (1971) 83–90.

L. Zhao, A model of limit order book dynamics and a consistent estimation procedure, Ph.D. thesis, Carnegie Mellon University, 2010.

I. M. Toke, "market making" in an order book model and its impact on the spread (2011) 49–64.

R. Cont, A. Kukanov, S. Stoikov, Stochastic models for order book dynamics, Operations research 58 (2010) 549–563.

L. Bauwens, N. Hautsch, Modelling financial high frequency data using point processes (2009) 953–979.

A. Ellul, C. W. Holden, P. Jain, R. Jennings, Determinants of order choice on the new york stock exchange, manuscript, Indiana University (2003).

A. Hall, N. Hautsch, Order aggressiveness and order book dynamics, Empirical Economics 30 (2006) 973–1005.

E. Bacry, M. Bompaire, S. Gaïffas, S. Poulsen, tick: a Python library for statistical learning, with a particular emphasis on time-dependent modeling, ArXiv e-prints (2017). arXiv:1707.03003.

H. Bodor, L. Carlier, Stylized facts and market microstructure: An in-depth exploration of German bond futures market, 2023. Working Paper.

F. Abergel, M. Anane, A. Chakraborti, A. Jedidi, I. M. Toke, Limit order books, Cambridge University Press, 2016.

A. Chakraborti, et al., Econophysics review: I. empirical facts, Quantitative Finance 11 (2011) 991–1012.

R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance 1 (2001) 223–236.

J.-P. Bouchaud, J. Bonart, J. Donier, et al., Trades, quotes and prices: Financial markets under the microscope, Cambridge University Press, 2018.

Appendix A. Complementary Stylized Facts Results

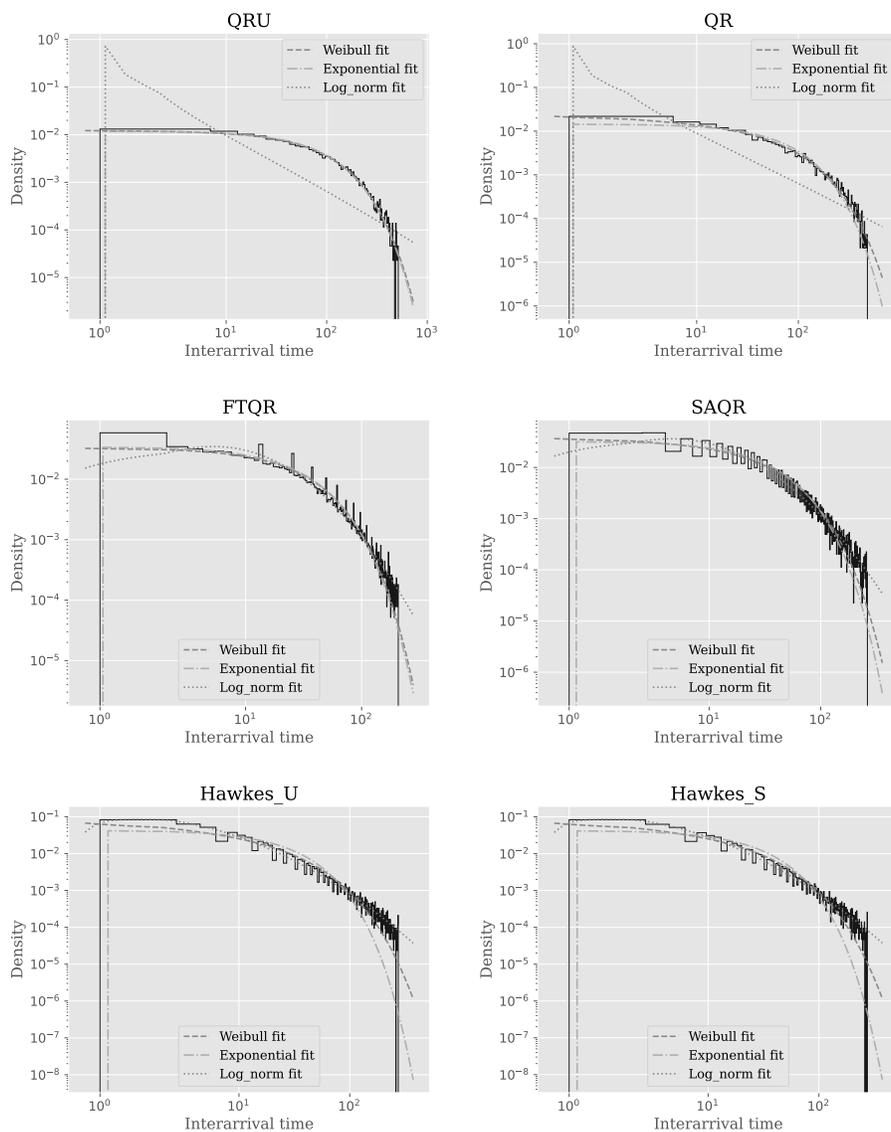*Appendix A.1. Weibull Fit of Interarrival Time of Trades*

Figure A.15: Distribution of interarrival times of market orders.

Another known stylized fact is the Weibull fit of interarrival times of market orders. Results in Figure A.15 show that all the models verify this stylized fact, as the Weibull distribution is the closest fit.
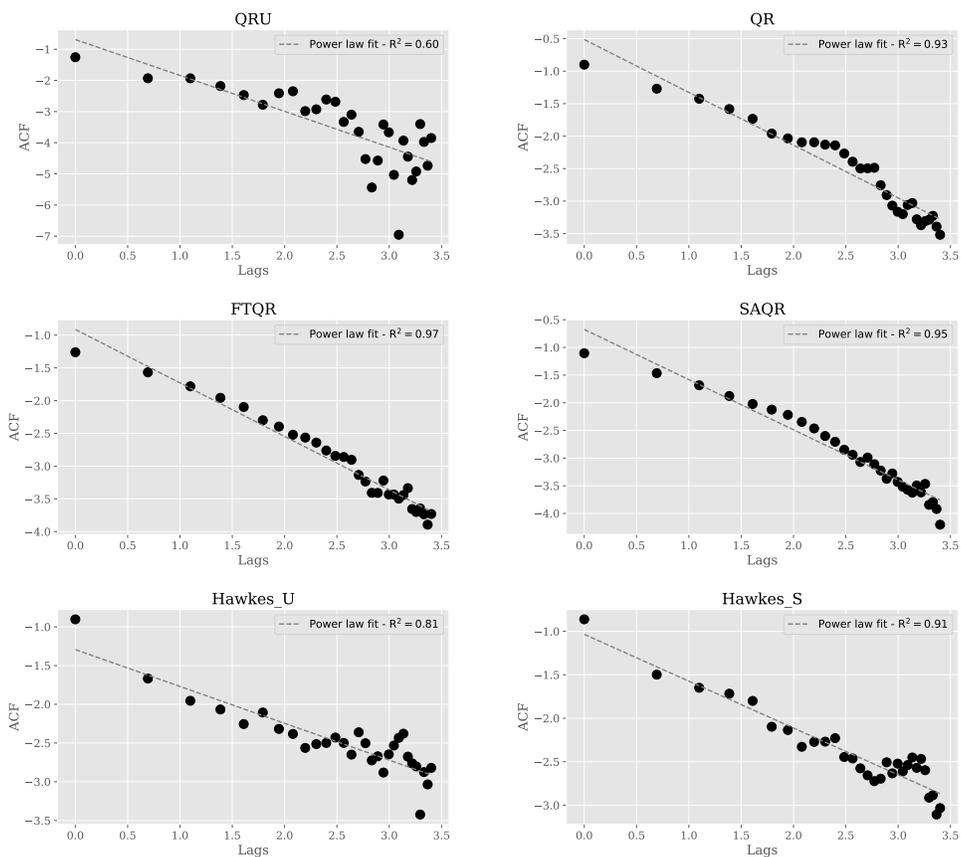
Appendix A.2. Long Range Dependence

Figure A.16: Long range dependence.

While a realistic market should show no correlation between returns, it is commonly observed in different markets that the correlation of absolute returns shows a slow decay (Cont, 2001). Moreover, this decay follows a power-law distribution. Figure A.16 shows how simulated markets exhibit this slow decay of absolute returns correlation, with a pretty good fit of power-law distribution, except for the QRU and Hawkes_U models.
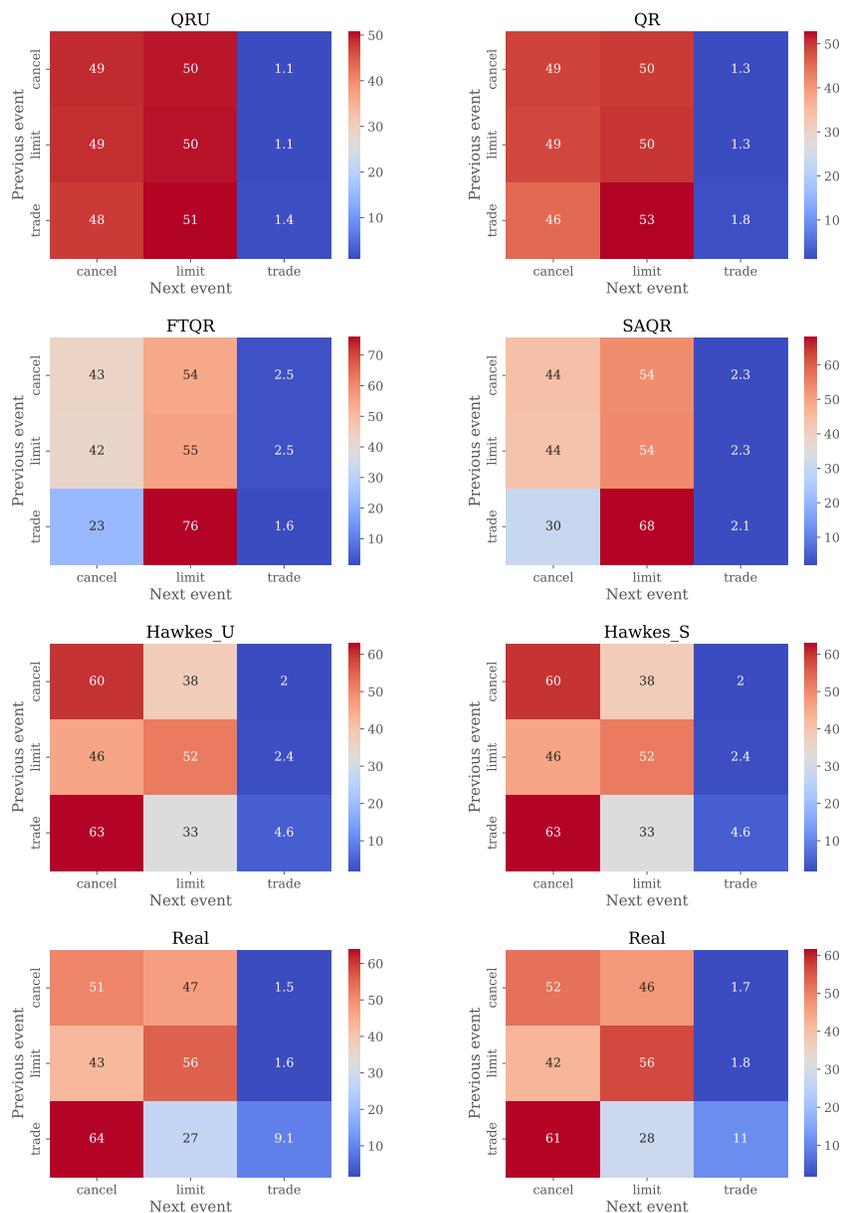
Appendix A.3. Excitation between events

Figure A.17: Transition probabilities between events.

Figure A.17 presents the matrix of conditional probabilities of an event E_{t^+} (columns) given a past event E_{t^-} (rows), denoted as $\mathbb{P}\left(E_{t^+}|E_{t^-}\right)$. The matrices reveal that the queue-reactive models replicate the empirical matrices quite inadequately, as evidenced by nearly uniform rows in the matrix. This indicates that the occurrence of a past event does not significantly influence the likelihood of future events in these models. In contrast, Hawkes-based models demonstrate a notable accuracy in reproducing the excitation between events, attributable to their design and the inclusion of a memory component.
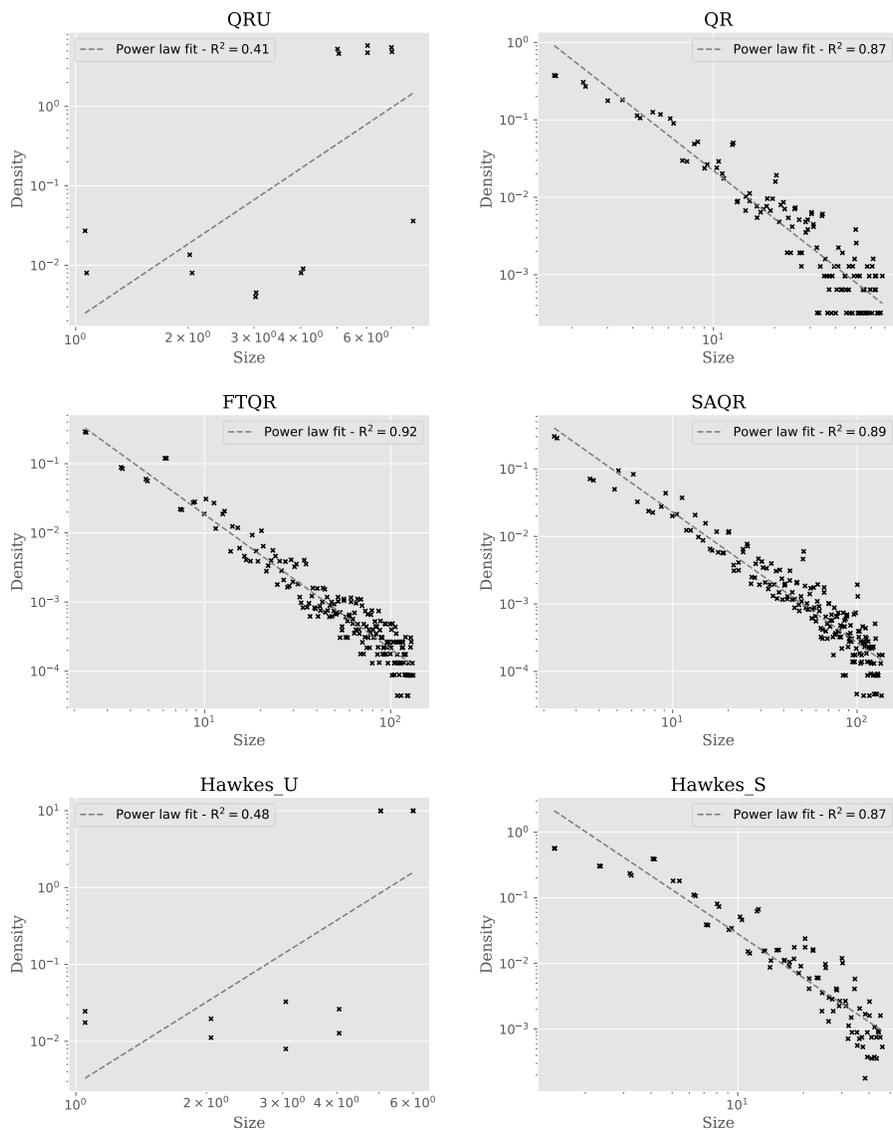
Appendix A.4. Power-law of Order Sizes

Figure A.18: Distribution of trade order sizes and power-law fit.

The power-law fit of order sizes is also one of the stylized facts that is verified in multiple markets (Chakraborti et al., 2011; Abergel et al., 2016; Bouchaud et al., 2018). In particular, Bund futures market exhibits this behavior (Bodor and Carlier, 2023). Figure A.18 shows that all the models satisfy this stylized fact, except for the QRU and Hawkes_U models, mainly because of the unitary nature of the order sizes.
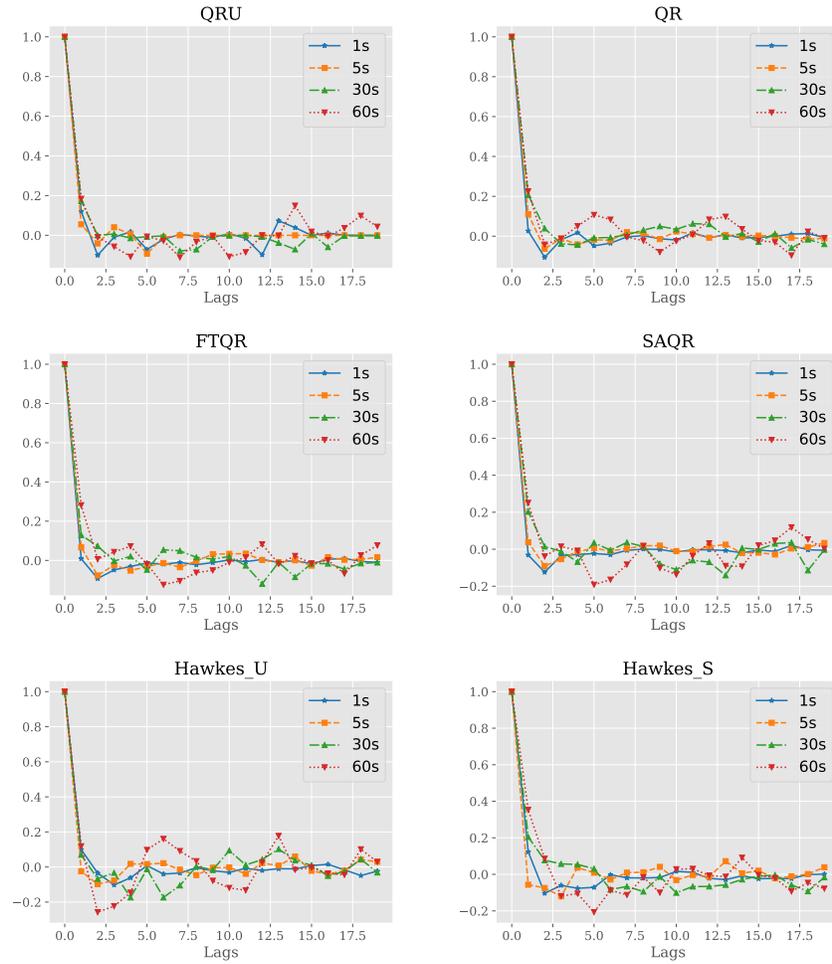
Figure A.19: Autocorrelation of returns for different sampling frequencies.

Figure A.19 displays the autocorrelation of returns for different sampling frequencies and for the six models, it shows that there is generally no autocorrelation between returns at different time scales.