Aalto University
School of Science

Master's Programme in Mathematics and Operations Research

Market Making in Intraday Power Markets

Ahti Korhonen

Master's Thesis
2025

Aalto University
School of Science

Author Ahti Korhonen

Title Market Making in Intraday Power Markets

Degree programme Mathematics and Operations Research

Major Systems and Operations Research

Supervisor Prof. Nuutti Hyvonen

Advisor Robert Soczewica (MSc.)

Collaborative partner Danske Commodities AS

Date 14^{th} December 2025 **Number of pages** 57 **Language** English

Abstract

The rapid growth of renewable energy production has led to a substantial increase in trading volumes on European intraday power markets, where continuous balancing of supply and demand is required since electricity cannot be stored at scale. Renewable generation is highly volatile and requires frequent adjustments through intraday trades to maintain system balance. Trading in the continuous intraday power markets is organized through a limit order book, which enables a market microstructure perspective on liquidity formation and trading behavior.

This thesis studies the market microstructure of the German continuous intraday market and analyzes the determinants of its liquidity. We extend existing research by examining the impact of intraday auctions on the continuous market and evaluate the market impact of aggressive buy- and sell-orders. Based on these empirical findings, we develop a market making model that incorporates observed microstructural features and resolves several inconsistencies in the existing market making literature.

The market making problem is formulated as a stochastic impulse control problem, leading to a Hamilton–Jacobi–Bellman quasi-variational inequality that can be solved numerically. Finally, we present an example for the German intraday market and show the resulting optimal quoting strategy for a market maker. The numerical example shows that our model is able to capture key microstructural features that match the observed data.

Keywords Market making, intraday power markets, market microstructure, stochastic optimal control, limit order book, inventory management

Aalto University
School of Science

Tekijä Ahti Korhonen

Työn nimi Markkinatakaus päivänsisäisillä energiamarkkinoilla

Koulutusohjelma Mathematics and Operations Research

Pääaine Systems and Operations Research

Työn valvoja Prof. Nuutti Hyvönen

Työn ohjaaja Robert Soczewica (MSc.)

Yhteistyötaho Danske Commodities AS

Päivämäärä 14. Joulukuuta 2025 **Sivumäärä** 57 **Kieli** englanti

Tiivistelmä

Uusiutuvan energian tuotannon voimakas kasvu on johtanut merkittävästi lisääntyneisiin kaupankäynti volyymeihin Eurooppalaisilla päivänsisäisillä sähkömarkkinoilla, joissa kysynnän ja tarjonnan tulee olla tasapainossa, sillä sähköä ei voida varastoida skaalautuvasti. Uusiutuvan energian tuotanto on hyvin vaihtelevaa ja vaatii jatkuvia korjauksia ennusteisiin. Tälläisia korjauksia tehdään päivänsisäisillä markkinoilla, jotta sähköverkko pysyy tasapainossa. Kaupankäynti jatkuvalla päiväsisäisellä markkinalla tehdään tarjouskirjan kautta, mikä mahdollistaa likviditeetin muodostumisen ja kaupankäynnin tutkimisen markkinan mikrostruktuurin näkökulmasta.

Tässä diplomityössä tutkitaan Saksalaisen jatkuvan päivänsisäisen markkinan mikrostruktuuria ja analysoidaan sen likviditeetin määrääviä tekijöitä. Laajennamme olemmassaolevaa tutkimusta selvittämällä päivänsisäisten huutokauppojen vaikutusta jatkuvaan päivänsisäiseen markkinaan sekä tutkimalla aggressiivisten osto- ja myyntitarjousten markkinavaikutusta. Kehitämme empiiristen havaintojen pohjalta markkinatakausmallin, joka korjaa monia aiempien tutkimusten epäjohdonmukaisuuksia.

Markkinatakausmalli muotoillaan stokastisena impullsikontrolliongelmana, josta voimme johtaa kvasivariationaalisen Hamilton-Jacobi-Bellman yhtälön, joka voidaan puolestaan ratkaista numeerisesti. Lopuksi tutkimme esimerkkiä liittyen Saksalaiseen päivänsisäiseen markkinaan ja johdamme optimaalisen tarjousstrategian markkinatakaajalle. Numeerinen esimerkki osoittaa, että malli johtaa tuloksiin jotka ovat linjassa havaitun data kanssa.

Avainsanat Markkinatakaus, päivänsisäiset sähkömarkkinat, markkinoiden mikrostruktuuri, stokastinen saato, tarjouskirja, inventaarion hallinta

Preface

Thanks to my family, friends, colleagues at DC, professors and staff at Aalto and the Finnish tax payers to helping me get through my studies and writing of this thesis.

Aarhus, Denmark, 14^{th} December 2025

Ahti Korhonen

Contents

List of Symbols

t	point in time
S_t^b	bid price processes
S_t^a	ask price process
S_t	spread process
B_t	market makers cash process
Q_t	market makers inventory process
R_t^b	market makers limit order bid regime
R_t^a	market makers limit order ask regime
N_i	counting process of order type i
$\lambda_i(t)$	arrival rate of order type i at time t
ν_i	baseline intensity of order type i
ψ_i	exponential coefficient of the intensity of order type i
ω_i	exponential growth rate of the intensity of order type i
τ_i	arrival time of order type i
v_i	volume of order type i
κ_i	shape parameter of volume distribution for order type i
χ_i	scale parameter of volume distribution for order type i
π_i	price of order type i
ξ	price change in ticks caused by order arrival or cancellation
ζ	volume of a market order sent by the market maker
δ	tick size of the traded asset
ρ	the market makers participation rate
$V(t, x_t)$	market makers value function at time t and state x_t
α	discount factor for the market makers limit order additions
β	discount factor for the market makers market orders
θ	market makers running inventory penalty
ϵ	rebate paid by the exchange for liquidity provisioning
η	exchange fee paid for liquidity taking

List of Figures

Abbreviations

IDA	Intraday Auction
OTC	Over-the-Counter
EPEX	European Power Exchange
SDAC	Single Day-ahead Coupling
SDIC	Single Intraday Coupling
TSO	Transmission System Operator
BRP	Balance Responsible Party
LOB	Limit Order Book
MO	Market Order
LO	Limit Order
RFQ	Request-For-Quote
AS	Avellaneda-Stoikov
HJB	Hamilton-Jacobi-Bellman
HJBQVI	Hamilton-Jacobi-Bellman Quasi-Variational Inequality

1 Introduction

From 2014 to 2024, the share of power generated from renewable energy sources in Europe increased from 17% to 47% [1]. Unlike conventional power generation, the amount of power produced by renewable energy sources is purely weather dependent and thus uncertain, requiring continuous updates to forecasts and trading positions. This updating is done on the intraday spot market. The largest European intraday spot market is operated by the European Power Exchange (EPEX SPOT). The volume traded on the EPEX SPOT intraday continuous market rose by about 600% from 31 TWh in 2014 to 215 TWh in 2024 [2]. The intraday continuous market is a key enabler in the green transition from fossil fuels to renewable energy sources, as it allows market participants to trade on shorter time scales leading to smaller forecast errors as well as allowing flexible energy applications like virtual power plants and energy storage systems to operate. Compared to other power markets, the trading in intraday continuous market is organized in a limit order book. As the liquidity grows, it becomes interesting how different aspects from more mature limit order book markets, like the equity or futures markets, are reflected in the intraday continuous market.

Since the trading in the intraday continuous markets is facilitated by a limit order book, a natural way to study its functioning is through the lens of market microstructure. Market microstructure as a research area is the study of price formation and the underlying mechanics at the smallest level of granularity. Understanding of market microstructure is of essence for modern market participants in helping to understand their trading costs and market impact as well as developing sophisticated trading strategies. While the integration of renewable energy sources has been a hot topic for many years, the research on the market microstructure of the intraday continuous market is still sparse and as the organization of the market keeps changing rapidly, many of these studies are already outdated.

Market making is a trading strategy where a market participant provides liquidity to the market by either quoting passive limit orders on both sides of the limit order book or giving prices to customers at which they are willing to buy and sell an asset. The market makers profits come from the difference between the price at which they buy and sell the assets, this difference is known as the bid-ask spread. Some exchanges also pay a commission called a rebate for liquidity provisioning. Market making is possibly a highly profitable strategy in the intraday power market given the large spreads if the market maker is able to fade adverse price movements and manage their inventory. Moreover market making enhances the overall market quality since added liquidity results in smaller spreads and less volatile prices.

The objective of this thesis is to study different aspects of the limit order book on EPEX SPOT intraday continuous market in Germany to gain a deeper understanding of the market microstructure, and on the basis of these findings develop a market making model tailored to its unique characteristics.

This thesis is structured as follows: In Section 2 we provide detailed descriptions of modern electronic trading and limit order book mechanics, organization and development of European power markets and the theory behind optimal market making. In Section 3 we conduct empirical analysis on the market microstructure of the German intraday continuous market and extend the existing research on the topic by assessing the market impact of single orders and the effect of the introduction of intraday auctions (IDAs) on different aspects of market quality. Section 4 develops the mathematical framework for market making in a limit order book in the intraday context. Section 5 presents the numerical results of our model and Section 6 concludes the thesis.

2 Background

2.1 Power Markets

Liberalization of power markets began in Chile in the early 1980s as a group of young economists nicknamed the 'Chicago Boys' brought back new pro-business ideas after studying in the United States [3]. The main motivation behind the liberalization was to promote technical innovation and efficiency. Before this, the generation, transmission and delivery of power was managed by state owned monopolies. After the reorganization of the markets in Chile, multiple countries soon followed by opening their power markets for competition with the UK in 1989 and the Nordic countries in the early 1990s. As a consequence of the liberalization, power prices today are determined on the basis of supply and demand. To facilitate price discovery, power exchanges were established. At their inception, the main role of a power exchange was to gather bids for demand and production of power and to determine the market clearing price based on them. After a series of regulatory changes, aimed primarily at increasing the integration of renewables and increasing liquidity through market coupling, the power markets today are much more complex. Power trading in most European countries, with some regional differences, follows the European Union's target model [4] of structuring the market into four individual markets, which can be separated by the timescale at which they operate. Figure 2.1 presents these four markets in chronological order as we get closer to delivery.

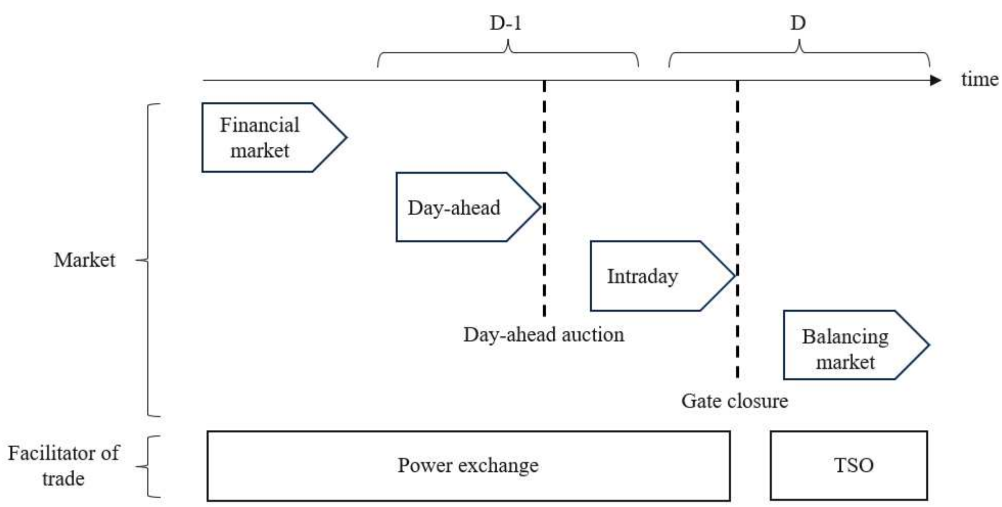

Figure 2.1: Timeline of power trading for a single delivery hour.

The financial market is quite different from the other three markets shown in Figure 2.1 as it almost never includes physical delivery of power, and the trading in the financial market is done purely for hedging long-term market exposures or speculating on price movements. Trading in the financial markets takes place either on a power exchange such as the European Energy Exchange (EEX) or Over-the-Counter (OTC).

The contracts traded on the financial market are different kinds of derivatives, most often futures.

2.1.1 The Day-Ahead Market

The day-ahead market, also known as the spot market, is the largest market of the four when measured by volume traded [2]. In the day-ahead market the prices are determined through a centralized auction that takes place the day before delivery (D-1) at 12:00 CET for all other European bidding zones except Switzerland and Great Britain. Most countries are divided into multiple bidding zones for geographical reasons or capacity constraints, for example, Denmark is divided into two bidding zones DK1 and DK2[1].

In the day-ahead auction, market participants submit so-called bidding curves, which state how much power they are able to generate or willing to consume at each price tick between the minimum and maximum prices at each hour[2] of the day. The price at which producers sell the power they generate is not arbitrary, but according to regulations, it must be priced at the marginal cost of production. The least expensive generation method is usually renewables and the most expensive methods are coal and gas plants. After the auction closes, the generation bids are organized in a merit order where the participants willing to produce power at the lowest prices are given priority. The organized bids form the supply curve for that given delivery hour. Similarly the bids for consumption are organized in a decreasing price order, meaning that participants willing to pay the highest price are given priority. The combined supply and demand curves are illustrated in Figure 2.2. In the simplified case that there would only be one bidding zone or no transfer capacity between bidding zones, the market clearing price could now be found at intersection of the supply and demand curves, and all producers would end up paying or receiving that same amount. The market clearing price would also be the marginal cost of the last production method that is turned on in order to meet the demand from the consumers.

In reality things are of course not so simple, European day-ahead markets are coupled through the Single Day-Ahead Coupling (SDAC) mechanism which combines multiple bidding zones in order to align supply and demand across bidding zones by optimizing cross-border flows. The exact pricing between the coupled zones is determined by the EUPHEMIA algorithm [6] which aims to maximize the total welfare of the system. Incase there is no capacity to flow power between bidding zones, the market clearing price will be different between the zones and if there is available capacity then the prices are the same.

Once the bids for production and consumption have been matched, the resulting trades are legally binding financial commitments within each bidding zone. Market participants are expected to deliver or consume the contracted volumes, but if their forecasts change they can adjust their position in the continuous intraday market. Any remaining deviations are settled in the balancing market at imbalance prices. This

[1]The Energy Island Bornholm is sometimes referred to as DK3, but for official trading on the exchange there exists only DK1 and DK2.

[2]The length of this interval may vary between countries, but one hour is the most common.

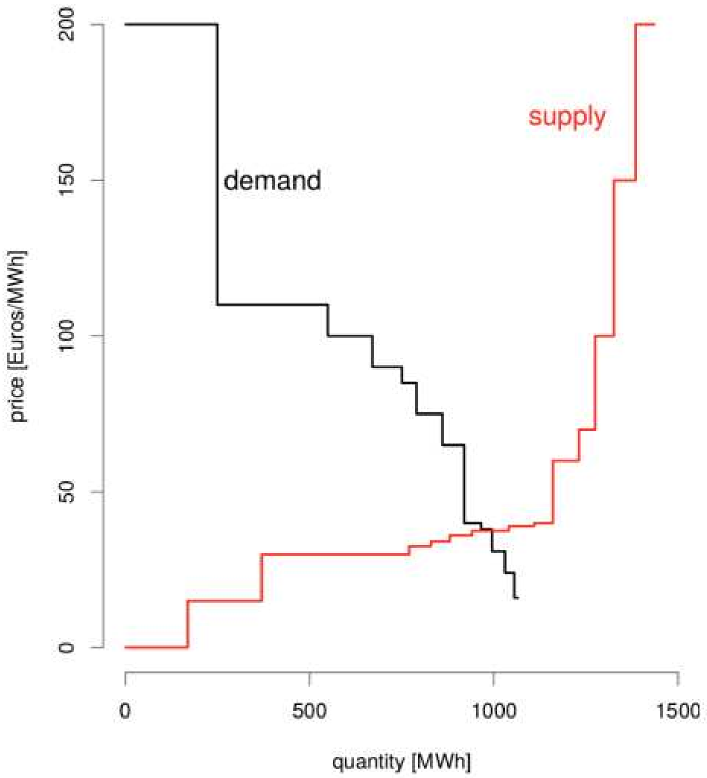

Figure 2.2: Combined supply and demand curves, the figure is from [5].

structure provides the transmission system operator (TSO) with a reliable baseline schedule, which is essential for maintaining system balance.

2.1.2 The Intraday Market

Trading on the intraday continuous market starts at 15:00 (D-1) CET after the day-ahead auction has closed and the spot prices have been published, and goes on until 5 minutes before delivery. Participants can trade contracts with hourly, half-hourly and quarterly deliveries. Trading on the intraday continuous market is conducted in a limit order book organized by the power exchange, the mechanism is similar to that of a stock exchange. In addition to the continuous market, participants can also trade in two or three, depending on the time of delivery, intraday auctions (IDAs) as of June, 2024 [7]. The IDAs work similarly as the day-ahead auction and provide participants with reference prices along the trading period and liquidity pooling as the intraday continuous market has historically had low liquidity. The IDAs are scheduled as follows:

- **IDA1**: closes at 15:00 (D-1) CET.
- **IDA2**: closes at 22:00 (D-1) CET.

- **IDA3**: closes at 10:00 (D) CET (available for products with delivery between 12:00-24:00 (D) CET).

Similarly as the day-ahead market, the intraday markets between different delivery zones are coupled by the Single Intraday Coupling (SIDC) mechanism to allow for cross-border flows. For IDAs, the SIDC works exactly the same as the SDAC, participants submit bids and the EUPHEMIA algorithm calculates a clearing price that maximizes the social welfare across delivery zones. On the intraday continuous market, market coupling is handled by a system called XBID [8], which keeps track of all the cross-border capacities made available by the TSOs and creates a shared limit order book where buy and sell orders between different delivery zones can be matched against one another, given that there is capacity between the zones. XBID adds complexity to the matching mechanism (see Section 2.2.1), as it must continuously track transmission capacities in real time and it also creates interesting implications for the market microstructure as we will see in Section 3. During the IDAs (± 20 min around the auction) and one hour before delivery cross-border matching via XBID is stopped, meaning that only orders in the same delivery-area can be matched against each other. Figure 2.3 illustrates the timeline of intraday trading (continuous market and IDAs) for a single delivery hour. As the penetration of renewables keeps growing, as is expected given EUs target of climate neutrality by 2050, the importance of the intraday continuous market keeps growing.

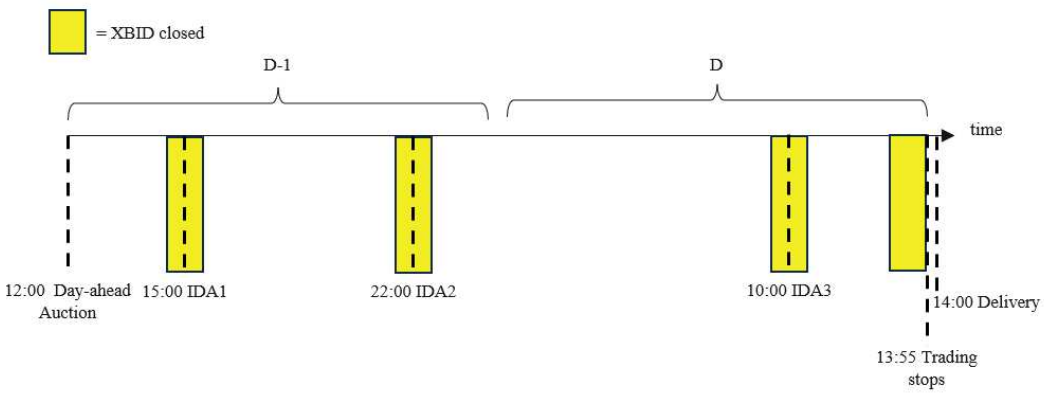

Figure 2.3: Timeline for trading on the intraday market for 14:00 delivery.

2.1.3 The Balancing Market

The balancing market is the final stage of physical power trading after the day-ahead and intraday markets. Its function is to keep supply and demand in balance in real-time during the delivery period. Unlike the financial, day-ahead, and intraday markets, which are operated by the power exchange, the balancing market is operated by the TSO, since they have the responsibility to keep the grid stable by ensuring that the frequency and voltage remain within acceptable limits. Most power grids in Europe are operated at 50 Hz frequency, and even a small disturbance of 0.5 Hz can cause a blackout. Since the balancing markets are operated by country-specific TSOs[3] there is more deviation between the way the markets are operated. Balancing markets can be operated on the basis of forecasts or reactively, i.e. when imbalances actually occur. Below is a broad description of how a TSO operates their balancing market in a reactive manner.

- **Reserve capacity procurement**: the TSO procures balancing capacity through daily auctions from balance responsible parties (BRPs) which are able to meet the technical requirements of the different balancing products. Capacities are procured for three main products that differ in terms of minimum bid sizes, activation, duration and ramp rates. The products are:
    1. Frequency Containment Reserves (FCR): the primary source of capacity for shorts bursts of power which can be activated in seconds.
    2. automatic Frequency Restoration Reserves (aFRR): The secondary mechanism for imbalance control which is activated automatically in 5 minutes. The European aFRR reserves are coupled through the PICASSO [9] mechanism.
    3. manual Frequency Restoration Reserves (mFRR): the market for balancing the largest deviations which can occur, for example, in the case of an unexpected outage of a large power plant. The mFFR is activated manually and thus takes the longest time to activate and can provide power for the longest. The European mFRR reserves are coupled through the MARI [10] mechanism.
- **Activation of balancing capacity**: When the system frequency deviates from 50 Hz, the TSO activates reserves either by up-regulation (increasing generation or reducing demand) or down-regulation (reducing generation or increasing demand).
- **Imbalance pricing**: To determine the price for the balancing power, the TSO creates a merit order, similar to the day-ahead auction, from the submitted bids that where activated during the delivery period and the price is then determined by the most expensive generation method. The BRPs that helped the system by providing capacity then receive this price from the TSO and the BRPs that contributed to the imbalance end up paying this price to the TSO.

If the TSO operates the balancing market based on imbalance forecasts, the mechanism is exactly the same as above with the distinction that the reserve capacities are activated when the forecasts show that an imbalance will occur instead when they actually occur.

[3] Most European countries have a single TSO except Germany which has 4 and Austria which has 2.

2.2 Limit Order Books

Today, a large portion of the trading of financial assets is done electronically. The most widely spread way to facilitate electronic trading is through a limit order book (LOB) [11]. Examples of limit order book markets include stock exchanges like the New York Stock Exchange (NYSE), NASDAQ and the London Stock Exchange (LSE). Other alternatives to facilitating electronic trading include request-for-quote (RFQ) systems, dark pools, auctions and automated market makers (AMMs). A market where trading is done in a LOB can also be referred to as an order driven market.

2.2.1 Description

Limit order book provides a snapshot to the current state of demand and supply for a certain asset. In the most basic setup, trading in a LOB is conducted using two types of orders: Market Orders (MOs) and Limit Orders (LOs). MOs are considered aggressive order types as they are executed immediately at the best possible price for an amount specified by the sender of the order or until there is no more liquidity in the LOB. LOs are considered passive orders as the sender specifies a quantity and a price at which they are willing to buy (sell) the asset, the specified price is usually lower (higher) than the current best ask (bid) which results in the order not getting immediately executed but getting placed in the LOB to wait for incoming MOs. MOs and LOs with an aggressive price that leads to immediate execution can be categorized as 'marketable orders'.

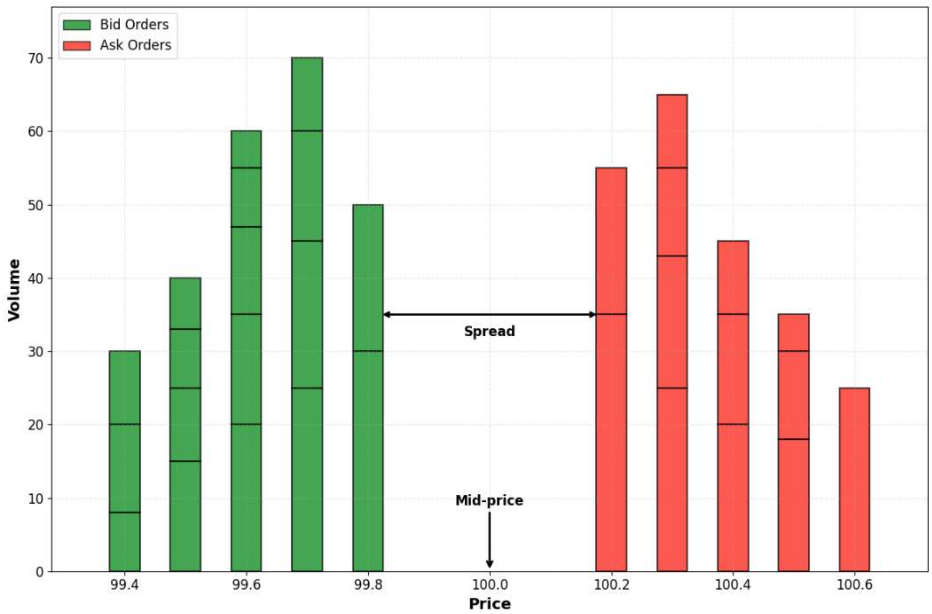

Figure 2.4: Limit order book snapshot.

Figure 2.4 presents a snapshot of a LOB. On the left, the green bars display buy orders, while on the right, the red bars display sell orders. Each block outlined with black represents an individual order, and the most competitive quotes are those at the top of each side: the best bid on the buy side and the best ask on the sell side. The difference between these two prices defines the bid–ask spread, while their average gives the mid-price, often used as a reference for the asset's current value. Trading is constrained by the tick-size, which is the minimum increment between price levels, that varies across assets depending on regulation, price and liquidity. Finally, it is important to note that the visible book reflects only part of the true market state: traders may conceal their intentions through mechanisms such as iceberg orders (see Section 2.2.3), which make the apparent depth of the book differ from its actual liquidity.

2.2.2 Matching Mechanism

To match incoming marketable orders with resting LOs, a rule-based mechanism is implemented by the trading venue. A common matching mechanism is the price-time priority rule, which is also known as the FIFO mechanism (first in, first out). Under the price-time priority rule, resting LOs are sorted by price in descending order for the bid side and ascending order for the ask side, meaning that buy (sell) orders with the highest (lowest) price are given the highest priority. Inside price levels, individual orders are sorted based on their arrival times, meaning that if a buy order arrives at the best bid which already has resting LOs, it will only be executed against an incoming marketable order once the orders before it are either executed or canceled. The price-time priority rule is used in, for example, equity markets as well as the intraday continuous market, which will be our main focus in the upcoming sections. For a detailed explanation on LOB mechanics under the price-time priority rule, see [12].

The pro-rata mechanism is also a matching mechanism worth mentioning as an alternative to the price-time priority rule. Under the pro-rata mechanism, orders are executed proportional to their volume relative to the total volume at that price level. The pro-rata thus prioritizes participants based on the liquidity that they provide instead of their speed. The pro-rata mechanism is used in many futures and options markets, such as the Chicago Board Options Exchange (CBOE).

2.2.3 Continuous Intraday LOB

Due to the nature of power as a commodity and how it is generated and consumed, the difference to other order driven markets is that the contracts traded exist only for a single trading session, meaning that hourly products with delivery at 14:00 today and 14:00 tomorrow have no connection to each other. Each contract is traded in a separate LOB, meaning that for a 15-minute product there is 96 independent LOBs open at the same time for a given day.

As discussed in the previous section the two basic order types are MOs and LOs. Since market dynamics have gotten more complex overtime, exchanges have introduced new order types which can cater to traders with different types of objectives. Most of the different order types can be derived from LOs and MOs. Below is a list of order types (excluding LOs and MOs) available for a trader on EPEX SPOT continuous intraday market:

- **Fill-or-Kill (FOK) order**: a market order that must be executed immediately and fully or not at all i.e. if the LOB does not contain enough liquidity to fully fill the order, the order is canceled and no trade happens.
- **All-or-Nothing (AON) order**: same as a fill-or-kill order, but without the urgency meaning that the order can rest in the LOB and wait until there is enough liquidity to fully fill the order up to the specified amount.
- **Immediate-or-Cancel (IOC) order**: an order that is executed immediately and then canceled for any unfilled portion of the order. The difference between an IOC order and a market order is that IOC orders are usually submitted with a price cap to limit risk.
- **Iceberg order**: a limit order that conceals the true volume of the order. When submitting the order, the trader specifies a price, volume and a visible quantity. The visible quantity is the part of the order that is shown in the LOB, and the order gets 'refilled' whenever the visible part is depleted. A marketable order can also match against the hidden part of the iceberg, given that its volume exceeds the visible part of the iceberg.

Once an order has been submitted, the trader has the following three actions available, given that the order is placed in the LOB:

- **Cancel**: the order is removed from the LOB and it ceases to exist.
- **Modify**: the trader can modify the price or quantity of the order, and in the case of an iceberg order, the visible quantity. When the price or quantity is modified, the order loses priority in the queue.
- **Deactivate/Activate**: the trader can also choose to 'hibernate' i.e. deactivate the order. In the case of hibernation the order is not visible in the LOB and can not be matched against. The reason why someone would like to hibernate their order could be, for example, in the case of losing a connection to the exchange.

2.3 Market Making

Before the introduction of electronic trading, trading was conducted physically in pits where designated market makers[4] received orders to buy and sell assets through an open outcry mechanism in which communication was done by shouting or using hand signals. With the advent of electronic trading and the introduction of limit order books, the role of market makers has changed significantly. Exchanges still employ designated market makers, but besides that, anyone can be a market maker simply by posting limit orders on both sides of the book and earning the spread. The business model in market making is very simple, the market maker posts quotes on both sides of the book and earns the bid-ask spread (plus-minus possible costs and rebates paid to/by the exchange) for each round trip trade. Making money is of course never risk-free, for market makers the risks stem from two sources:

- **Inventory risk**: In order for the market maker to complete a round trip of trades she should be able to buy a certain amount of an asset and then sell the assets at a higher price. Ideally, for this to happen the arrival rates of buy and sell MOs should be identical. In reality however, this is not often the case which leads to price fluctuations in the asset. For example, if the market maker holds a large negative inventory and the price moves up, the value of the market makers inventory takes a negative hit.

- **Adverse selection risk**: In essence market makers are always giving away free-options when they are posting LOs. If a market participant has more knowledge about the assets future value, they can exploit the free option given by the market maker who ends up buying/selling at an unfavourable price. A market participant with information about the future value of the asset can be referred to as an informed trader.

These risk can be accounted for by modeling order flow dynamics and posting quotes that take into account the market makers current inventory and possible adverse selection risks. The rest of this thesis will focus on managing inventory risks, for an example on managing adverse selection risk, see [13].

For further context it is also important to distinguish between market making in quote-driven and order-driven markets. As already discussed, market making in order-driven markets can be done by virtually anyone with access to the exchange that is maintaining the LOB. In quote-driven markets, market making is done by specialist firms, usually large investment banks, that provide quotes for customers at their request through RFQ-systems. Exaples of quote-driven markets include fixed-income and FX markets. This distinction is important for modeling purposes that will be explained in the subsequent sections.

[4]These were people/companies that were working for the exchange and had the obligation to provide liquidity to an asset for which they were rewarded for.

2.3.1 Avellaneda-Stoikov Model

In their seminal paper in 2008 Avellaneda and Stoikov [14] introduced an inventory management model for market making that is based on the theory of stochastic optimal control, the model will be referred to as the AS model from here on. The AS model is based on earlier work by Ho and Stoll in 1985 [15] where they present a model for inventory management for a monopolistic market maker. In the AS model, the goal is to maximize the expected utility of the market makers risk-adjusted profit and loss at the end of a trading session by continuously posting buy and sell limit orders to the market. The market makers value function is given by

$$u(x, s, q, t) = \mathbb{E}_t[-\exp(-\gamma(X_t + q_t S_t))],$$

from which we obtain the optimization problem

$$u(x, s, q, t) = \max_{\delta^a, \delta^b} \mathbb{E}_t[-\exp(-\gamma(X_T + q_T S_T))].$$

The maximization of the expression above is done by controlling the spread parameters δ^b and δ^a of the limit orders, meaning that the market maker will post buy orders at a price of $p^b = s - \delta^b$ and sell orders at a price of $p^a = s + \delta^a$, continuously updating the spread parameters based on the current inventory and market dynamics. S denotes the mid-price process of the asset that is given by

$$dS_t = \sigma dW_t,$$

where dW_t is standard one dimensional Brownian-motion and σ is a constant which denotes the assets volatility. X_t is the cash process of the market maker that is given by

$$dX_t = p^a dN_t^a - p^b dN_t^b,$$

where dN_t^a and dN_t^b are independent Poisson processes with intensities $\lambda^a(\delta^a)$ and $\lambda^b(\delta^b)$, denoting the number of sold and bought assets respectively. Naturally the market makers inventory at time t is then given by

$$q_t = N_t^b - N_t^a.$$

By applying results from dynamic optimization and stochastic calculus, it can be shown that the market makers value function solves the following Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} &u_t + \tfrac{1}{2}\sigma^2 u_{ss} \\ &+ \max_{\delta^b} \lambda^b(\delta^b)\left[u(s, x - s + \delta^b, q + 1, t) - u(s, x, q, t)\right] \\ &+ \max_{\delta^a} \lambda^a(\delta^a)\left[u(s, x + s + \delta^a, q - 1, t) - u(s, x, q, t)\right] = 0 \end{aligned}$$

with a terminal condition

$$u(s, x, q, T) = -\exp(-\gamma(x + qs)).$$

The HJB equation can then be simplified using an ansantz of the form

$$u(s,x,q,t) = -\exp(-\gamma x)\exp(-\gamma\theta(s,q,t)), \tag{1}$$

which reduces our HJB equation to

$$\begin{aligned}
&\theta_t + \tfrac{1}{2}\sigma^2\theta_{ss} - \tfrac{1}{2}\sigma^2\gamma\theta_s^2 \\
&+ \max_{\delta^b}\left[\frac{\lambda^b(\delta^b)}{\gamma}[1 - e^{\gamma(s-\delta^b-r^b)}]\right] \\
&+ \max_{\delta^a}\left[\frac{\lambda^a(\delta^a)}{\gamma}[1 - e^{-\gamma(s+\delta^a-r^a)}]\right] = 0
\end{aligned} \tag{2}$$

and the terminal condition to

$$\theta(s,q,T) = qs,$$

where r^b and r^a are the reservation prices i.e. the prices at which the market maker is insignificant to buying or selling one asset given their utility function. These prices can be defined as:

$$u(x - r^b(s,q,t), s, q+1, t) = u(x,s,q,t)$$

and

$$u(x + r^a(s,q,t), s, q-1, t) = u(x,s,q,t).$$

By applying the definitions of the reservation prices to the ansantz (1), we find that

$$r^b(s,q,t) = \theta(s,q+1,t) - \theta(s,q,t) \tag{3}$$

and

$$r^a(s,q,t) = \theta(s,q-1,t) - \theta(s,q,t) \tag{4}$$

depend directly on the function θ. Given the reservation prices (3) and (4), and applying them to our first order optimality condition (1), we obtain the optimal spreads δ^b and δ^a, which are given by the following relations

$$s - r^b(s,q,t) = \delta^b - \frac{1}{\gamma}\ln\left(1 - \gamma\frac{\lambda^b(\delta^b)}{(\partial\lambda^b/\partial\delta)(\delta^b)}\right)$$

and

$$r^a(s,q,t) - s = \delta^a - \frac{1}{\gamma}\ln\left(1 - \gamma\frac{\lambda^a(\delta^a)}{(\partial\lambda^a/\partial\delta)(\delta^a)}\right).$$

The model assumes that the arrival rate is an exponential function of the spread, which in the case of symmetric arrival rates can be written as $\lambda^a(\delta) = \lambda^b(\delta) = Ae^{-k\delta}$, where A and k are positive constants. The intuition is that the LO is more likely to get filled when it is closer to the best bid/ask, and the likelihood decreases exponentially as the spread increases. In order to obtain the optimal spread values, the reservation prices need to be solved first. This can be done by first solving (2) numerically and then plugging the θ values in to equations (3) and (4) for bid and ask prices, respectively.

2.3.2 Avellaneda-Stoikov Extensions and Inconsistencies

After 2008 a lot of the research related to market making has been related to extending the AS model while ignoring its inconsistencies with real market dynamics. The main issue with the AS model is the independence of the mid-price process and the order arrival processes, meaning that the mid-price can jump up at the same time as a large market sell order arrives which is impossible given real-world LOB mechanics. One attempt to fix this issue was introduced by Fodra and Pham 2015 [16]. They model the mid-price as

$$S_t = S_0 + \delta \sum_{T_n \leq t} J_n,$$

where S_0 is the starting mid-price, δ is the tick-size, $\{(T_n, J_n)\}$ is a Markov reneval process related to the order arrivals and the jump-directions. $T_n \in \mathbb{R}_+$ marks the order arrival times and $J_n \in \{-1, +1\}$ indicates whether the price has jumped up $(+1)$ or down (-1). Now, although the two processes are correlated, the relationship between them is still statistical, not causal, meaning that there is a non-zero probability that the price jumps up on a sell order arrival and vice-versa, so the price inconsistency still exists. Moreover, there is nothing that guarantees the order arrival and direction processes to jump at the same time.

Market makers goal is to always end the trading session with no inventory, meaning that during the trading session the market maker tries to avoid accumulating large positive or negative inventories. The AS model does not include penalization for terminal inventory or inventory accumulated during the trading session. Fodra and Labadie 2012 [17] tackled this challenge in their model which incorporates a penalty for non-zero terminal inventory in the value function. The market makers optimization problem in their model is stated as

$$u(t, s, q, x) = \max_{\delta^a, \delta^b} \Big[\mathbb{E}\big[-e^{-\gamma(X_T + Q_T S_T - \eta Q_T^2)}\big] \Big],$$

where η dictates how severely the terminal inventory is penalized.

The AS model shows how a market maker can find optimal quotes for a single asset, but in reality most maker makers are providing liquidity for multiple assets simultaneously. In 2013 Fodra and Labadie [18] introduced another extension to the AS model that takes into account multiple assets through a covariance matrix Λ of the asset prices. The market makers optimization problem in the multi-asset scenario is

$$u(t, s, q, x) = \max_{\delta^a, \delta^b} \Big[\mathbb{E}\big[X_T + Q_T S_T - \epsilon\big(\eta Q_T' \Omega_T Q_T + \nu \int_t^T Q_\xi' \Lambda_\xi Q_\xi d\xi\big)\big] \Big],$$

where Ω_T is a $M \times M$ matrix representing the cost of executing a portfolio of Q_T at the market at the end of the trading session. The model also includes penalization for running and terminal inventories.

Although most, if not all, of the AS model extensions are applied to high-frequency electronic trading platforms, which are order-driven, they lack many critical features of limit order books. Most notable missing features are:

- **Price-time priority**: In the AS model and its extensions it is assumed that the quotes can be updated continuously i.e. there is no cost in updating a quote. However, when quotes are updated all the time in a LOB, the market makers orders will always be at the end of the queue, meaning that they will hardly ever get filled in a liquid market. In a quote-driven market this assumption of continuous updates makes sense, since the update does not cost anything. A customer can request a price for an asset through, for example, a Bloomberg terminal and then ask for a new price one minute later and for the market maker there is no cost if the price given to the customer is now different than previously.

- **Price ticks**: Due to LOB mechanics, a price can only be a multiple of tick-size i.e. a jump process. In most models, the mid-price is modeled as a diffusion process which is an approximation at best and does not capture many of the stylistic facts of a LOB such as clustering of order arrivals. In addition, when a price is modeled as a continuous variable, the resulting prices from complex market making models might be irrelevant due to the tick-size constraint.

- **Execution probability**: In the AS model, the arrival rate of the MOs is given by $\lambda(\delta) = Ae^{-k\delta}$, which can also be thought of as a proxy for the fill rate of the market makers LOs. The spread δ is modeled as a continuous variable, but in reality asset price is always a multiple of the tick-size, meaning that the spread should be modeled as a step-function. Moreover, modeling the arrival rate as an exponential function makes an implicit assumption of the order book shape, however, for example, in very liquid markets like ETF or futures markets the probability that a large MO walks the book to the second price level is less than 3% [19]. In this case the spread based optimization becomes irrelevant and market makers optimization problem becomes to just decide when to be posted at the top of the book and when not.

- **Order size**: To simplify the model, the AS model assumes that the market maker posts one order on each side of the book and that the incoming MOs are all of the same size. In reality, market makers layer multiple quotes on each side of the book in order to have a strong queue position once the price levels ahead are depleted. Since MOs are of course not all of the same size, this kind of an assumption can create scenarios where the market makers inventory grows to unacceptable levels.

In the following sections, we will dive deeper in to the microstructure of the intraday continuous market and then develop a market making model that fixes previously mentioned inconsistencies in the AS model and incorporates microstructural features from our analysis.

3 Empirical Analysis and Quality of Intraday Power Markets

In order to successfully develop trading algorithms, it is crucial to understand who the counterparties are, what are the economics behind their trading decisions and how their decisions affect the market. The next sections contain empirical analysis of different aspects of trading in the intraday continuous market. We will look at price formation based on order arrivals and market quality measures such as bid-ask spread distribution, volatility and LOB depth distribution of the intraday power market. Market quality can be defined as how easily the market facilitates trading and can be measured in multiple dimensions such as liquidity, price efficiency and transparency.

The data used for the analysis is level 3 LOB data from Q2 2025 EPEX SPOT Germany, meaning that the data contains all limit order additions and cancellations as well as market orders for each one-hour product with delivery in one of the four German delivery zones. Due to the XBID mechanism, the data also includes orders from other delivery zones, but the origin of the orders is not included in the data. From the data we can reconstruct the state of the LOB at any point in time during the trading session and study how individual orders affect different aspects of the LOB. The tick size for EPEX SPOT continuous market is 0.01 €/MWh and minimum volume increment is 0.1 MW, and the minimum and maximum prices are −9999 €/MWh and 9999 €/MWh, respectively.

3.1 Intraday Volume

Figure 3.1 shows a heatmap and the median volume of executed marketable orders in one minute buckets averaged over all trading sessions in the dataset for hour 14 delivery. The volume is plotted on a log(1+volume) vs. time-to-delivery axis, where volume is in megawatts. The IDAs are highlighted in the plot. During the IDAs and the final hour of trading, XBID is closed, meaning that the LOB only contains orders with delivery at one of the four German delivery zones. The effect of this can be seen on the graph as the executed volume drops visibly during the IDAs and spikes right after when the LOBs are joined together and overlapping orders are executed.

Compared to other financial markets, the volume pattern is quite different. Usually, the volume pattern follows a U-shape [20] that can be explained by traders acting on new information during the open and close out positions as the day ends. In intraday power markets, the traded volume grows rapidly as the gate closure nears. Since the intraday market is mostly used for balancing out positions for renewable generation, the volume pattern is expected given that the forecasts for production become more accurate over the trading session. It can also be seen clearly that the volume picks up at around 6 to 7 hours before the gate closure which is the time when the morning shift starts for traders in Europe. Since the market is open 24/7, the trading is done in tree shifts: morning, evening and night shifts or alternatively, the night shift might be working from a country with a more suitable time-zone like Singapore. The traded products are divided between shifts based on the delivery hours and this of course also effects the volume pattern. For small market participants it might be economically infeasible to have traders working night and weekend shifts, which means that the volume pattern is also affected by the day of the week, with weekends having less traded volume [21].

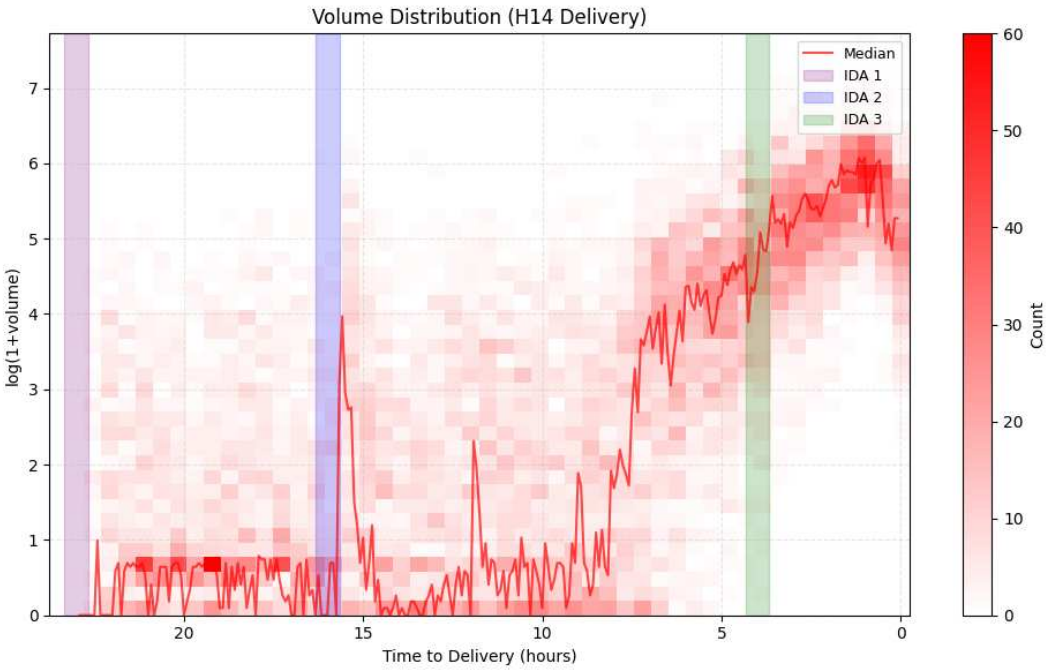

Figure 3.1: Volume distribution for hour 14 delivery.

3.2 Bid-ask Spread

Bid-ask spread, which we will from now on refer to just as spread, is the difference between the best ask and best bid $s = s_a - s_b$ in the LOB at a given time. The spread can be viewed as the cost of immediate execution as a trader can choose to execute their order passively by placing a LO in the LOB or aggressively by sending a MO which crosses the spread. The spread is also an important proxy for liquidity as it is an implicit transaction cost.

Figure 3.2 shows the average behavior of the spread during a trading session for hour 14 delivery. The plot shows the median spread in one minute buckets averaged over the whole dataset. Excluding the IDAs, the spread seems to follow an L-shaped curve which is in line with the findings by Balardy in [22]. The spread decreases with time as there is more activity closer to the delivery. During the IDAs the spread increases significantly which is expected given that the spread is negatively correlated to volume according to [23]. The median spread of the average trading session is 0.89 €/MWh, which is considerably less than 3.52 €/MWh which was the median spread in 2015 reported by Balardy in [22]. This decrease can be attributed to the annually increasing volume in the intraday market. Although the median spread is more than three times smaller than the median spread in 2015, it is still 89 times the tick size, which compared to equity markets is massive. In equity markets the spread is typically few times the tick size at most.

Figure 3.2: Bid-ask spread distribution as a function of time to delivery.

Comparing Figures 3.2 and 3.3 is interesting from the market makers perspective as it would make sense that the spread would increase in times of high volatility as the market makers have to price the future uncertainty in to their LOs, but this does not seem to happen in the intraday market. The weekend effect in liquidity can also be observed through the spread as Balardy reports in [22] that the spread is on average 13% higher during the weekends. The tick size also has a large effect on the spread distribution. Given that the tick size of 0.01 €/MWh is quite small compared to the average price per megawatt, the observed distribution is not that surprising. For a more in-depth discussion on the effects of tick size on the spread distribution, see [20].

3.3 Volatility

Volatility is another key dimension of market quality which measures fluctuations in prices. Volatility can be divided into two separate components, fundamental volatility and microstructure noise. Fundamental volatility measures the fluctuations in the assets true value, and microstructure noise measures fluctuations in price due to microstructural market events. In this thesis we do not make an explicit distinction between the two, but due to the time scale at which the volatility is calculated, the measures below lean more towards microstructure noise. Volatility can be measured in a number of ways, the simplest being the standard deviation of returns, also known as realized volatility. It is also important to distinguish between historical volatility, which is what we are looking at, and implied volatility. Implied volatility is an estimate of the future volatility of an asset that can be calculated from option prices.

Figure 3.3 visualizes the volatility of the German intraday market for hour 14 delivery defined in two ways. The volatility measure shown in the left panel of the figure shows the whole distribution as well as the median standard deviation of the log-returns in one-minute buckets. The issue with this definition is that the volatility is naturally higher in times where there are more trades, which in our case is the end of the trading session. The volatility measure on the right panel is defined as the maximum price minus the minimum price normalized by the median price in one-minute buckets. Using this definition, we are not affected by the frequency of the trades. Looking at the median plots, the pattern is quite similar for both definitions, extremely high volatility during the first IDA, almost no volatility apart from the second IDA and then the volatility starts to pick up close to the end of the trading session.

When comparing the volatility patterns with the volume pattern, they look highly correlated, which is expected given that volatility is the result of trading activity. From the balancing point of view, it would be expected that the volatility measure on the right would decrease closer to the delivery as the production and delivery forecasts become more accurate and thus the true value of the delivered power should become more accurate, but it looks like the microstructure noise dominates both measures. Using less granular time-frames could reveal patterns about the fundamental volatility, but from the market makers perspective the long-term behavior is not that interesting.

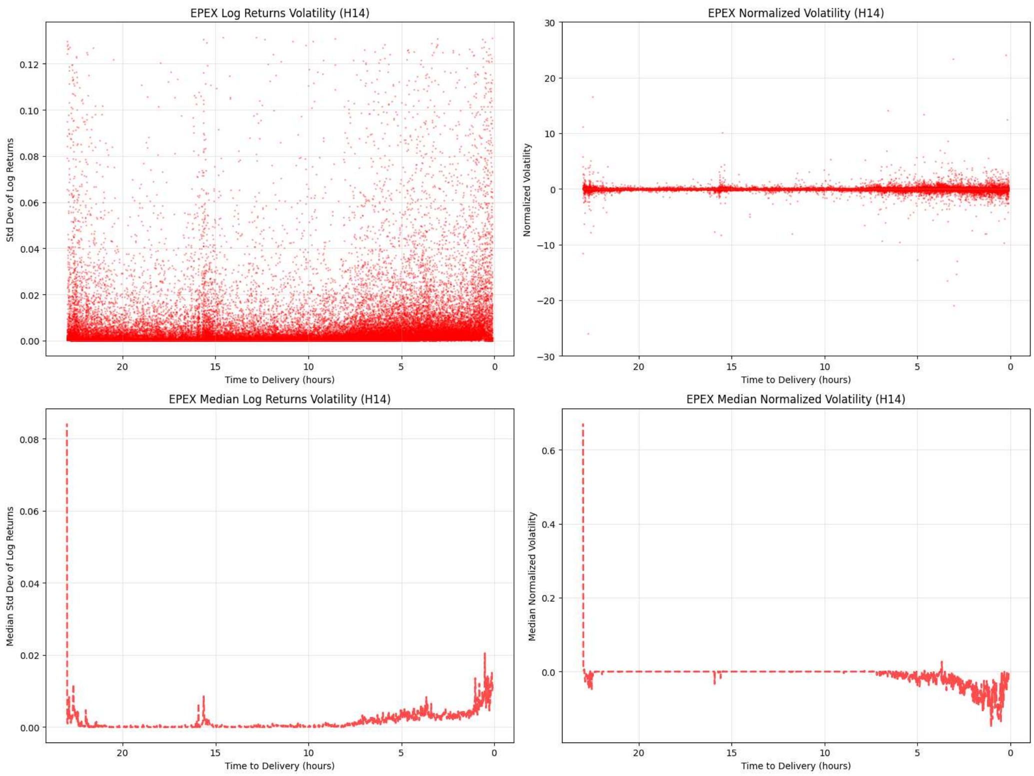

Figure 3.3: Volatility distribution as a function of time to delivery.

3.4 LOB Depth

The quality of the market is not only dependent on the incoming orders or transaction costs, but also on the available liquidity in the LOB. In this section we will have a look at the LOB depth of the German intraday market at the top price levels. Depth in this context refers to the available liquidity in the LOB at a given time. As market participants look to minimize their market impact, the depth of the LOB and size of the incoming orders are also highly correlated.

Figure 3.4 shows depth of the best bid and ask as a function of the time to delivery. The figure shows that the available liquidity on both sides is relatively similar and centered around 5 megawatts. 5 megawatts is also the minimum visible quantity for an iceberg order on EPEX SPOT, which might explain the high number of orders with a quantity of 5 megawatts. Interestingly, the volume on both sides drops visibly after the third IDA. This is most likely caused by execution algorithms getting more aggressive towards the end of the trading session as they split orders into small pieces and try to execute passively in order to get a better price. Verifying this is of course not possible since all order information is anonymous.

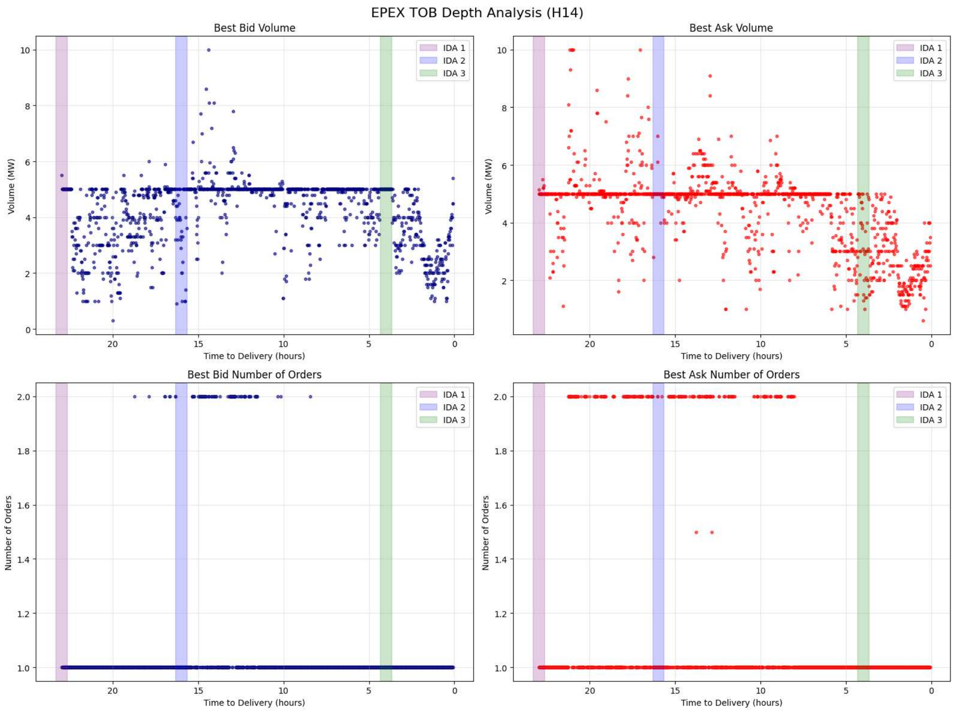

Figure 3.4: Top of the book depth as a function of time to delivery.

Both figures 3.4 and 3.5 show that the number of individual orders at each price level is almost always one or two, which means that the order queues are very weak. For market making purposes strong queues would be optimal since then the price impact and the resulting negative effect on the market makers inventory would be smaller. Figure 3.5, which shows the cumulative liquidity and the total number of the orders at the top five levels of both sides of the LOB, confirms the earlier observation that towards the end of the trading session the orders get smaller and the liquidity at the top levels of the book gets more scarce.

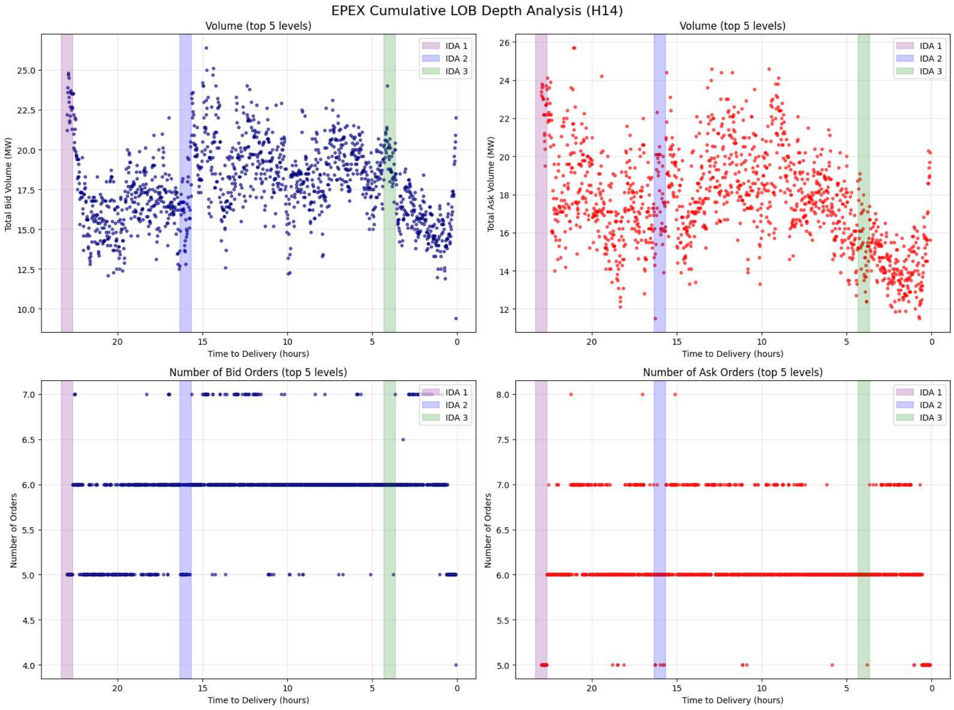

Figure 3.5: LOB depth for top 5 price levels as a function of time to delivery.

Figure 3.6 shows the mean and median distance between the top price levels over the average trading session. The distribution for both sides is very similar to the bid-ask distribution, excluding the IDAs. The distances get smaller towards delivery which would indicate that the total liquidity in the LOB increases. During the IDAs and the final hour of trading, the distances jump up as a natural result of the XBID being turned off.

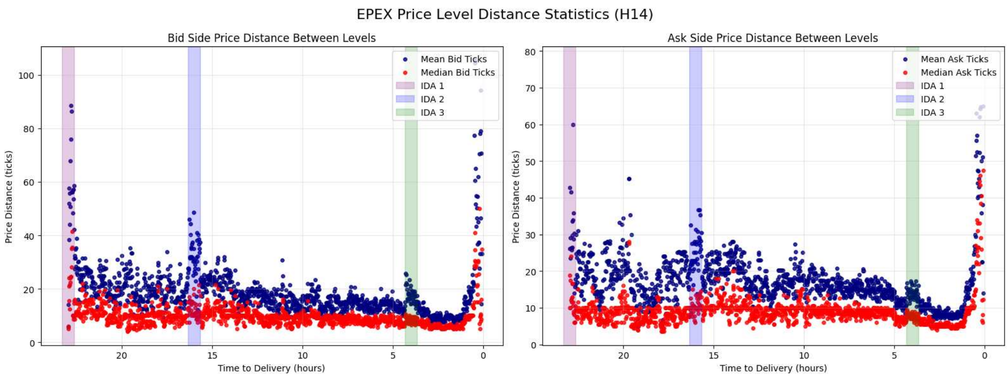

Figure 3.6: Mean and median distance between the top 5 price levels as a function of time to delivery.

3.5 Order Arrival rates and Cancellations

A determining factor in the evolution of the LOB is the arrival and cancellation rates of LOs and MOs. Modeling the arrival rates is also a key topic in the market makers optimization problem as the intensity of these processes determines the frequency of the LOs being lifted and the resulting gain from the spread.

Figure 3.7 shows the average intensity of MO arrivals during a trading session for hour 14 delivery. The distinction between aggressive and non-aggressive MOs is that aggressive MOs move the current mid-price i.e. the volume of the MO is large enough to deplete all the liquidity at the current best bid or ask, and non-aggressive MOs have less volume than is currently available at the best bid or ask. The intensity of MO arrivals is highly correlated with the volume pattern in Figure 3.1, but the pattern is smoother and follows an exponential shape. The intensity drops significantly only in the final hour of trading, and the IDAs seem to have little to no impact on the intensity. The share of aggressive MOs to non-aggressive MOs seems to be constant. Von luckner and Kiesel also provide strong empirical evidence about the clustering of MOs in the German intraday market in [24]. Clustering means that successive orders are more likely to be of the same sign, which can be attributed to market participants splitting their MOs into smaller pieces, following the behavior of other traders or acting on the same information in timely proximity.

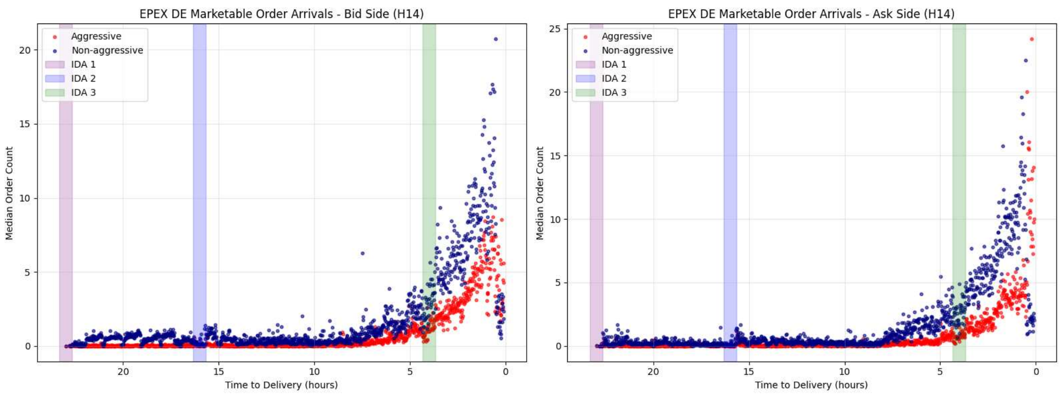

Figure 3.7: MO arrival distribution as a function of time to delivery.

The ability to cancel and place new limit orders is critical for the market maker in order to change their LO placement in the LOB upon arrival of new market events. Figures 3.8 and 3.9 present the intensities of LO arrivals and cancellations for the hour 14 delivery. Aggressive LO arrival is defined as a LO that arrives inside the spread thus changing the mid-price and non-aggressive arrival is an order that arrives at the current best bid or ask, or a worse price. Aggressive cancellation is defined as a canceled order that moves the mid-price which happens when the canceled order is the only order at the best bid or ask. The patterns for LO additions and cancellations are almost identical, which is not a coincidence: most of the orders are placed and then canceled in the next few seconds. The patterns for additions and cancels are not as smooth as the MO arrival pattern and are significantly affected by the IDAs. During the IDAs, the intensities for both additions and cancellations drop significantly and jump right after when XBID is turned on again. The drop in intensity in the final hour is also very clear. The share of aggressive LO additions and cancels is relatively constant with small spikes during the IDAs due to lower liquidity in the LOB.

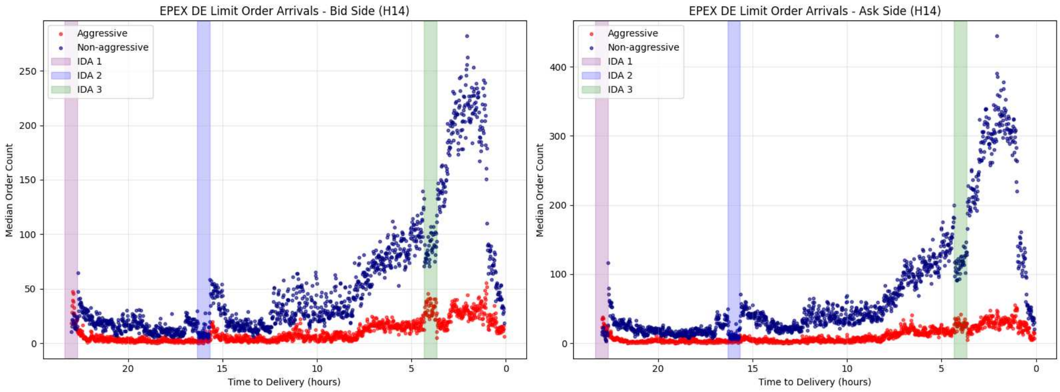

Figure 3.8: LO arrival distribution as a function of time to delivery.

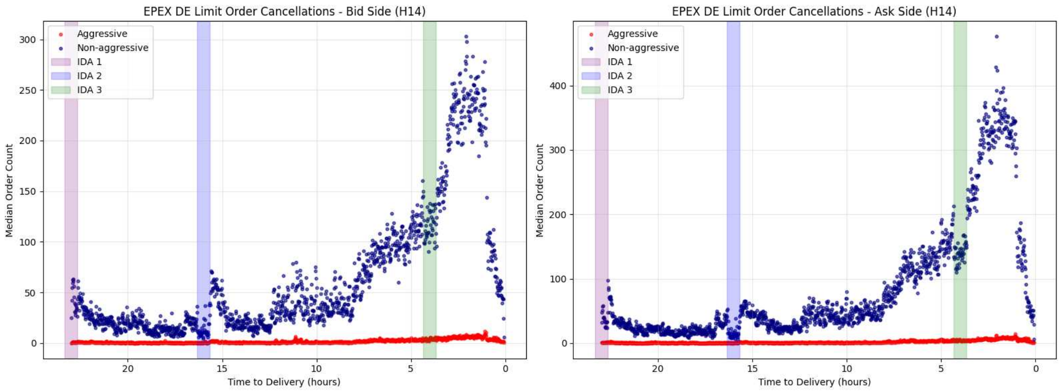

Figure 3.9: LO cancel distribution as a function of time to delivery.

3.6 Market Impact Function Approximation

Market impact measures the relationship between the volume of an incoming order and the resulting price change. The square-root law for price impact [25] is considered a universal truth in many asset classes. The law states that the order size and the resulting price difference have the following relationship:

$$\Delta p = c\sigma \left(\frac{Q}{V}\right)^{\delta},$$

with $\delta \approx 1/2$, where σ is the daily volatility of the asset, Q is the order size, V is the average daily volume and c is a coefficient that is fitted to the data. The coefficients c and Δp can be found by fitting a linear regression where the dependent variable is $\log(Q/V)$ and the independent variable is $\log(\Delta p/\sigma)$. In traditional financial literature, market impact is measured from metaorders, which are large orders that are split into smaller child orders. The volume Q is the volume of the metaorder i.e. the sum of the child order volumes, and the price difference is measured as the difference of the price before executing the first child order and the price after executing the final child order. Due to the lack of metaorder data, the market impact approximation presented in Figure 3.10 is calculated based on individual orders and can be treated as a proxy for the market impact.

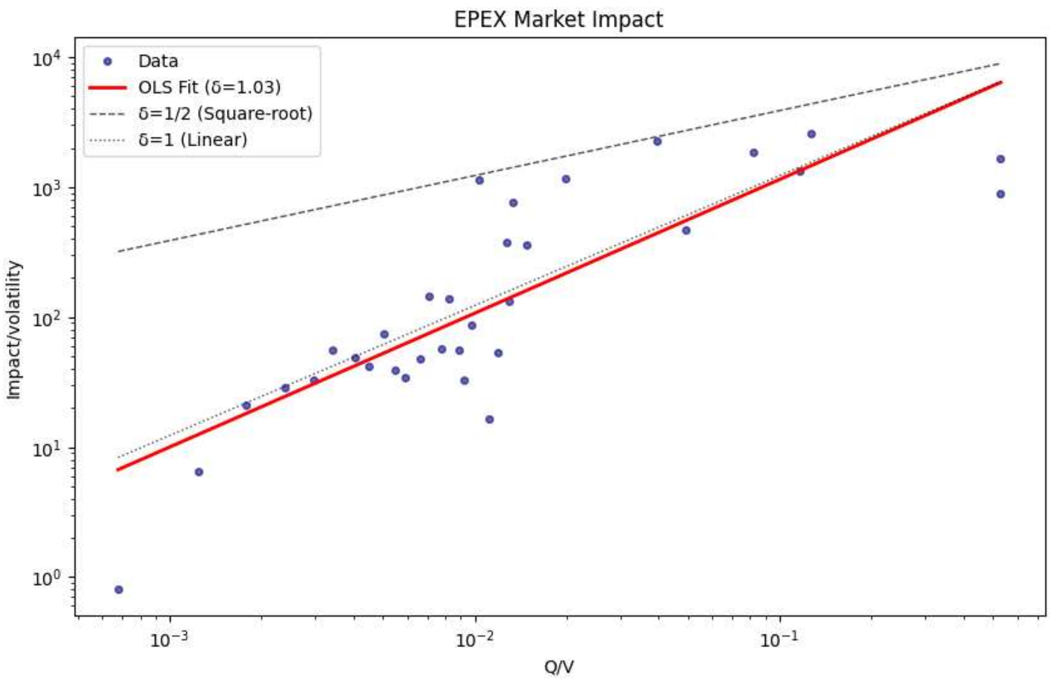

Figure 3.10: Approximation of the market impact of single MOs.

As the figure shows, the square-root law does not hold for the intraday power market since the exponent is $\delta \approx 1.03$ which is very close to linear. The exponent of the market impact gives us an understanding of the average shape of the LOB, with a bigger exponent implying lower liquidity. However, a linear model might be too simple in the case of intraday power markets. In [26] the authors model the market impact in intraday power markets using splines with two dimensions: time to delivery and order size. This kind of an approach may be more suitable since as we saw in Section 3.3, the liquidity at the top of the book is decreasing as delivery gets closer.

4 Market Making in Intraday Power Markets

This section presents the market making model based on stochastic optimal control which aims to solve the previously mentioned inconsistencies in the existing literature and incorporates microstructural features from our empirical analysis in order to make the model suitable for the intraday power market. More specifically, the goal is to model the market events that determine the evolution of the LOB to achieve a model that is consistent with the real-world LOB dynamics. Using the LOB model we solve the optimal quoting strategy for a high-frequency market maker in the German intraday power market. The model is largely based on the work of Law and Viens [27].

4.1 Consistent LOB Model

The state of the LOB at any point in time can be determined by the LOs and MOs that have arrived up until that time point. All other Features of the LOB, such as the spread and mid-price, can be derived from these past order arrivals. From here on, we will use the following nomenclature, $S_t^b, S_t^a, S_t^m = (S_t^b + S_t^a)/2, S_t = S_t^a - S_t^b$ to denote the bid-price, ask-price, mid-price and spread. $\tau_m^b, \tau_m^a, \tau_l^b, \tau_l^a, \tau_c^b, \tau_c^a$ denote the arrival times of MO, LOs and LO cancellations, and v and π denote the corresponding order volume and price of the arriving LO, respectively. In order for a LOB model to be consistent with the real-world LOB dynamics it must satisfy the following three rules

1. **Direction Consistency**:

   - On the arrival of a marketable buy (sell) order, only the ask (bid) price can move up (down), while the bid (ask) price has to stay unchanged:

$$\mathbb{P}\big(\{S^a_{\tau^a_m} \geq S^a_{\tau^a_m -}\} \cap \{S^b_{\tau^a_m} = S^b_{\tau^a_m -}\}\big) = \mathbb{P}\big(\{S^b_{\tau^b_m} \leq S^b_{\tau^b_m -}\} \cap \{S^a_{\tau^b_m} = S^a_{\tau^b_m -}\}\big) = 1.$$

   - On the arrival of limit buy (sell) order, only the bid (ask) price can move up (down) while the ask (bid) price has to stay unchanged, given that the LO arrives inside the spread. Otherwise, both the bid and ask prices will remain unchanged:

$$A_1 = \big\{S^b_{\tau^a_l} = S^b_{\tau^a_l -}\big\} \cap \big\{S^a_{\tau^a_l} < S^a_{\tau^a_l -}\big\} \cap \big\{\pi^a_l < S^a_{\tau^a_l -}\big\},$$

$$A_2 = \big\{S^b_{\tau^a_l} = S^b_{\tau^a_l -}\big\} \cap \big\{S^a_{\tau^a_l} = S^a_{\tau^a_l -}\big\} \cap \big\{\pi^a_l \geq S^a_{\tau^a_l -}\big\},$$

$$A_3 = \big\{S^a_{\tau^b_l} = S^a_{\tau^b_l -}\big\} \cap \big\{S^b_{\tau^b_l} > S^b_{\tau^b_l -}\big\} \cap \big\{\pi^b_l > S^b_{\tau^b_l -}\big\},$$

$$A_4 = \big\{S^a_{\tau^b_l} = S^a_{\tau^b_l -}\big\} \cap \big\{S^b_{\tau^b_l} = S^b_{\tau^b_l -}\big\} \cap \big\{\pi^b_l \leq S^b_{\tau^b_l -}\big\},$$

$$\mathbb{P}\big(A_1 \cup A_2\big) = \mathbb{P}\big(A_3 \cup A_4\big) = 1.$$

   - On the cancellation of a limit buy (sell) order, only the bid (ask) price can move down (up) if the canceled order is the only order at the best bid (ask), while the ask (bid) price will remain unchanged. Otherwise, both the bid and ask prices will remain unchanged:

$$B_1 = \left\{S^b_{\tau^a_c} = S^b_{\tau^{a-}_c}\right\}\cap\left\{S^a_{\tau^a_c} \geq S^a_{\tau^{a-}_c}\right\}\cap\left\{\pi^a_c = S^a_{\tau^{a-}_c}\right\},$$

$$B_2 = \left\{S^b_{\tau^a_c} = S^b_{\tau^{a-}_c}\right\}\cap\left\{S^a_{\tau^a_c} = S^a_{\tau^{a-}_c}\right\}\cap\left\{\pi^a_c > S^a_{\tau^{a-}_c}\right\},$$

$$B_3 = \left\{S^a_{\tau^b_c} = S^a_{\tau^{b-}_c}\right\}\cap\left\{S^b_{\tau^b_c} > S^b_{\tau^{b-}_c}\right\}\cap\left\{\pi^b_c > S^b_{\tau^{b-}_c}\right\},$$

$$B_4 = \left\{S^a_{\tau^b_c} = S^a_{\tau^{b-}_c}\right\}\cap\left\{S^b_{\tau^b_c} = S^b_{\tau^{b-}_c}\right\}\cap\left\{\pi^b_c \leq S^b_{\tau^{b-}_c}\right\},$$

$$\mathbb{P}\big(B_1 \cup B_2\big) = \mathbb{P}\big(B_3 \cup B_4\big) = 1.$$

2. **Timing Consistency**: the bid and ask prices can only move at times when orders arrive or are cancelled:

$$\mathbb{P}\big(\{S^b_t = S^b_{t-}\}\cap\{S^a_t = S^a_{t-}\} \mid t \notin \Gamma\big) = 1,$$

where Γ is the set of all arrival times of MOs, LOs and LO cancellations.

3. **Volume Consistency**:

- On the arrival of marketable buy (sell) order whose volume is larger than the available liquidity at current best ask (bid), the price moves up (down) and otherwise stays unchanged:

$$\mathbb{P}\Big(\big(\{S^a_{\tau^a_m} > S^a_{\tau^{a-}_m}\}\cap\{v^a_m \geq Q^a_{\tau^{a-}_m}\}\big)\cup\big(\{S^a_{\tau^a_m} = S^a_{\tau^{a-}_m}\}\cap\{v^a_m < Q^a_{\tau^{a-}_m}\}\big)\Big) = 1,$$

$$\mathbb{P}\Big(\big(\{S^b_{\tau^b_m} < S^b_{\tau^{b-}_m}\}\cap\{v^b_m \geq Q^b_{\tau^{b-}_m}\}\big)\cup\big(\{S^b_{\tau^b_m} = S^b_{\tau^{b-}_m}\}\cap\{v^b_m < Q^b_{\tau^{b-}_m}\}\big)\Big) = 1.$$

- If the volume of the canceled limit buy (sell) order is equal to the volume of volume available at the current best bid (ask), the bid (ask) price moves down (up), and otherwise the price stays unchanged:

$$\mathbb{P}\Big(\big(\{S^a_{\tau^a_c} > S^a_{\tau^{a-}_c}\}\cap\{v^a_c = Q^a_{\tau^{a-}_c}\}\big)\cup\big(\{S^a_{\tau^a_c} = S^a_{\tau^{a-}_c}\}\cap\{v^a_c < Q^a_{\tau^{a-}_c}\}\big)\Big) = 1,$$

$$\mathbb{P}\Big(\big(\{S^b_{\tau^b_c} < S^b_{\tau^{b-}_c}\}\cap\{v^b_c = Q^b_{\tau^{b-}_c}\}\big)\cup\big(\{S^b_{\tau^b_c} = S^b_{\tau^{b-}_c}\}\cap\{v^b_c < Q^b_{\tau^{b-}_c}\}\big)\Big) = 1.$$

One could argue that these rules only apply when no orders can arrive at the same exact moment, but given how the matching engine of the exchange operates the orders are always handled sequentially even in the case that they arrive on the same exact nanosecond. The conditions stated above are necessary in order to model the LOB realistically, regardless of any assumptions about the arrival rates or volume distributions. However, in practice achieving a consistent LOB model is extremely difficult since we would need to keep track of the depths of all the price levels on both sides. In the next section we present the weakly consistent LOB model which only models the best bid and ask queues, and is thus only direction and timing consistent.

4.2 Weakly Consistent LOB Model

The behavior of the best bid and ask queues is determined by the orders arriving at the top of the book, these orders can be divided into two categories based on their aggressiveness. An aggressive order is one which moves the current mid-price i.e. changes either the price of the best bid or best ask, and non-aggressive orders don't affect the mid-price. Table 1 shows how the arriving MOs and LOs, and cancelled LO can be categorized by their aggressiveness. The $+/-/0$ sign indicates how the corresponding price is affected by the given order type.

Type	Order Arrival Event	Bid Price	Ask Price
1	aggressive market buy	0	+
2	aggressive market sell	-	0
3	aggressive limit buy	+	0
4	aggressive limit sell	0	-
5	aggressive limit buy cancellation	-	0
6	aggressive limit sell cancellation	0	+
7	non-aggressive market buy	0	0
8	non-aggressive market sell	0	0
9	non-aggressive limit buy	0	0
10	non-aggressive limit sell	0	0
11	non-aggressive limit buy cancellation	0	0
12	non-aggressive limit sell cancellation	0	0

Table 1: Order Classification

Based on this categorization we can form the following simple multivariate point process $N(t) = (N_1(t), \dots, N_{12}(t))$ which denotes the number of order arrivals for each order type up to time t. Using these processes, we can form the following relationships for the bid and ask prices:

$$S^b(t) = S^b(0) + (N_3(t) - N_2(t) - N_5(t))\delta$$

and

$$S^a(t) = S^a(0) + (N_1(t) + N_6(t) - N_4(t))\delta.$$

The bid price is simply the initial price S_0 plus the total number of price changes in ticks caused by the incoming/canceled orders, and the logic for the ask price is exactly the same.

In addition to jump times $\tau_n \in \mathbb{R}_+ = (0, \infty)$, each order arrival process has random marks related to the price impact and order volume. The price impact is denoted by $\xi_i \in \mathbb{N}_0$ and marks the number of ticks that the bid or ask price moves after an order arrival, for non-aggressive orders $\xi_i = 0$ always. The volume of the incoming order is denoted by $v_i \in \mathbb{R}_+$. The multivariate point process thus becomes $N_i(t) = (dt \times d\xi \times dv)$, and has a compensator of the form $\lambda_i(t)\mu_i(t, dv \times d\xi)dt$

where $\lambda_i(t)$ is the intensity of the ground process $N_i(dt \times \mathbb{R}_+ \times \mathbb{N})$ and $\mu_i(t, dv \times d\xi)$ is the conditional volume and jump distribution. From these it follows that the bid price is

$$S_t^b = s^b + \int_{(t_0,t]\times\mathbb{Z}_+} \xi\delta(N_3(t) - N_2(t) - N_5(t))(dt \times d\xi),$$

the ask price is

$$S_t^a = s^a + \int_{(t_0,t]\times\mathbb{Z}_+} \xi\delta(N_1(t) + N_6(t) - N_4(t))(dt \times d\xi),$$

and the spread is given by

$$S_t = s + \int_{(t_0,t]\times\mathbb{Z}_+} \xi\delta(N_1 + N_2 - N_3 - N_4 + N_5 + N_6)(dt \times d\xi).$$

In order to avoid a situation where the spread would became less than δ, we impose that the intensities of the aggressive LO additions are conditional on the spread based on the following relationship, where $\mathbb{1}()$ is the indicator function:

$$\lambda_i(t) = \lambda_i'(t)\mathbb{1}(S(t) > \delta), \quad i = 3, 4.$$

So when the spread is equal to one tick, the intensity of aggressive LO additions drops to zero. Now, it is clear that our model is direction and timing consistent, but since we do not keep track of the depths of the price levels, the model is not volume consistent.

4.2.1 Order Arrival Intensities and Mark Distributions

In theory the order arrival intensities can be any predictable non-negative stochastic processes. To exactly match the order arrival behavior of the German intraday power market, a Hawkes process would be optimal as per [24], but this would make the later introduced numerical scheme intractable. Because of this, we will settle for an exponentially increasing intensity of the form $\lambda_i(t) = \nu_i + \psi_i e^{\omega_i t}$, where ν_i is the baseline intensity of order type i, ψ_i is the exponential coefficient of order type i and ω_i is the exponential growth rate of the order type i. And for order types $i = 3, 4$, the intensity is of the form $\lambda_i(t) = (\nu_i + \psi_i e^{\omega_i t})\mathbb{1}(S(t) > \delta)$, $i = 3, 4$.

A practical way to find the arrival rate parameters is to divide historical data from multiple trading days in to one minute buckets based on the time to delivery and use non-linear least squares to fit the function $\lambda_i(t)$ to the data. Figure 4.1 shows the intensity functions fitted to historical data. The data is the same as in Section 3.5 but limited to the time window between IDA2 and IDA3. The data is not divided into bid and ask orders, but we can see that $\lambda_i(t)$ fits the average historical behavior quite well with an average R-squared value of $R^2 = 0.74$.

For the order volume distribution and size of the price jump distribution any distributional assumptions will be compatible with our model. Usually in financial literature it is assumed that order volume follows a power-law distribution. For the power market there is no existing research on the topic, but a Weibull-distribution

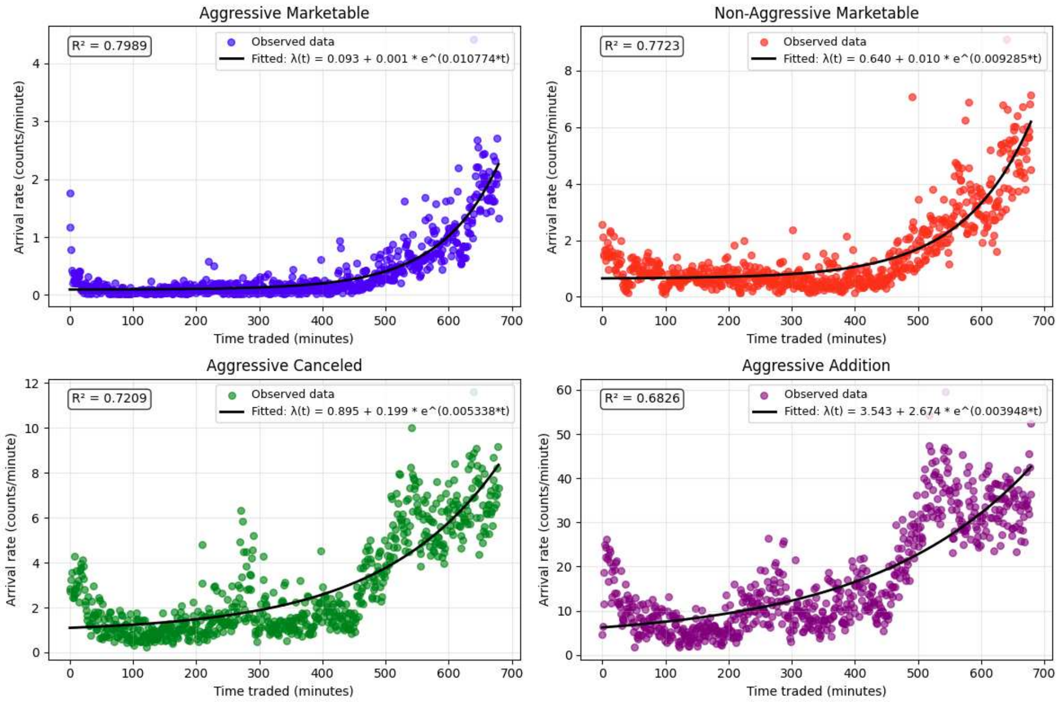

Figure 4.1: $\lambda_i(t)$ fitted to historical order arrival data.

seems to fit historical data the best for both aggressive and non-aggressive MOs. Figure 4.2 shows the cumulative probability distributions for both of these. The weibull-distribution parameters for the aggressive MOs are $\kappa = 0.662$ and $\chi = 5.15$, and for the non-aggressive MO distribution the parameters are $\kappa = 0.629$ and $\chi = 1.33$.

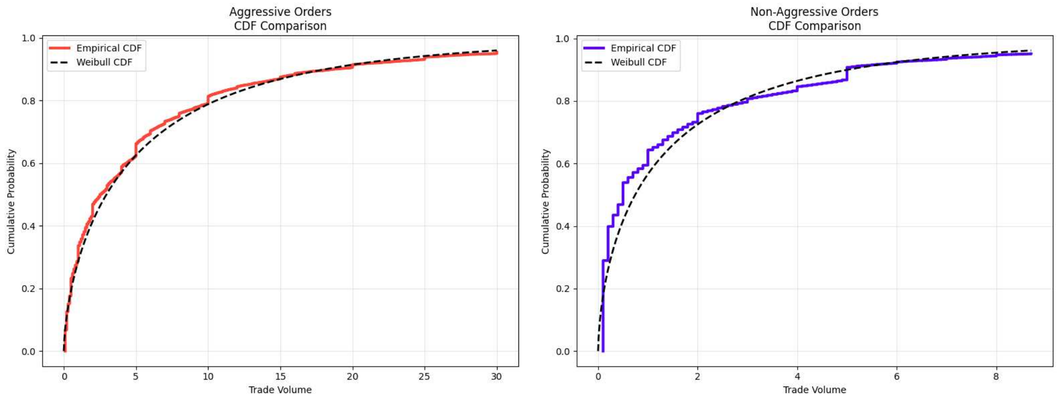

Figure 4.2: Weibull-distribution fitted to historical order volume data.

The price jump distribution should be discrete and non-negative given the tick-size restriction. The challenge with the intraday power market is the low liquidity which leads to a very large distribution in the jump sizes. To tackle this, the jump size should be conditional on the order volume and tied to the market impact function for MOs. However, for simplicity, in the first iteration of the model the distributions are assumed to be independent and the tick size jump distribution is more restricted than the true distribution.

The joint distribution of volume and jump size is given by

$$f^i_{v,\xi}(v,\xi) = f^i_v(v) f^i_\xi(\xi),$$

where

$$f^i_v(v;\kappa_i,\chi_i) = \frac{\kappa_i}{\chi_i}\left(\frac{v}{\chi_i}\right)^{\kappa_i-1} e^{-(v/\chi_i)^{\kappa_i}}$$

for both aggressive and non-aggressive MOs and

$$f^i_\xi(k) = \frac{\sum_{j=1}^{N_i} \mathbb{1}(\xi_j = k)}{N_i},$$

which is just the empirical distribution of the price jump for each order type. For the numerical example presented in Section 5, we have made the following assumptions of the price jump distributions while acknowledging that the real distribution is much wider. For aggressive MOs the distribution is $\mathbb{P}(\xi = 1) = 0.5$, $\mathbb{P}(\xi = 2) = 0.25$, $\mathbb{P}(\xi = 3) = 0.15$, $\mathbb{P}(\xi = 4) = 0.09$ and $\mathbb{P}(\xi = 5) = 0.01$, for aggressive LO additions $\mathbb{P}(\xi = 1) = 1$ and for aggressive LO cancellations $\mathbb{P}(\xi = 1) = 0.67$, $\mathbb{P}(\xi = 2) = 0.22$ and $\mathbb{P}(\xi = 3) = 0.11$.

4.3 Market Making Model

In the following sections we will develop the model for market making in the German intraday power market. The market maker is allowed to only quote top of the book and it can be assumed that the dynamics of the best bid and ask queues follow the ones presented earlier.

The market maker can choose to post LOs at the top of the book on both sides, only one side or withdraw the currently active quotes from the LOB from one side or both. These regimes are denoted by $R_t^b \in \{0, 1\}$ and $R_t^a \in \{0, 1\}$, where $R_t^b = 1$, $R_t^a = 1$ would mean that the market maker has quotes placed on both sides of the LOB at time t. In the case that either R_t^b or R_t^a is zero at $t - 1$, then the market maker will place a LO at the top of the book on the respective side, and if both R_t^b and R_t^a equal to one at $t - 1$, then the market maker does not need to do anything but let system fluctuate according to its dynamics. Additionally, the market maker can choose to send a MO of size $\zeta \in \mathbb{R}$ to hedge excess inventory. In total the market maker has three decisions to make on each time step: whether to change the current LO regime for either side, whether to send a MO to get rid of excess inventory, and in case that the market maker chooses to hedge, how should the MO be sized.

4.3.1 The Market Makers Optimization Problem

The market makers goal is to maximize the expected total wealth (cash + inventory) at the end of a trading session by finding the set of optimal controls $u \in \mathbb{U}$, where $\mathbb{U}$ is the set of admissable controls consisting of a sequence of decision variables $\{(\tau_n, r_n^b, r_n^a, \zeta_n)\}_{n\geq 1}$. The market maker is also subject to a running inventory penalty for holding non-zero inventory through out the trading session.

The market makers cash position evolves according to the following process

$$\begin{aligned}
B_t = b_{t_0} &+ \int_{(t_0,t]\times\mathbb{R}_+} \rho v R_{r^-}^a (S_{r^-}^a + \epsilon)(N_1 + N_7)(dr \times dv) \\
&- \int_{(t_0,t]\times\mathbb{R}_+} \rho v R_{r^-}^b (S_{r^-}^b - \epsilon)(N_2 + N_8)(dr \times dv) \\
&- \sum_{\tau_n \in (t_0,t]} \big((S_{\tau_n^-}^a + \eta)\zeta_n^+ - (S_{\tau_n^+}^b - \eta)\zeta_n^-\big).
\end{aligned}$$

The first integral term describes the money received from MOs lifting our ask quotes plus the rebate $\epsilon \geq 0$ paid by the exchange for liquidity provision. The $S_r^a + \epsilon$ term is the money received, but since the market maker is not necessarily always active in the LOB, we have to account for this through the $R_{r^-}^a \in \{0, 1\}$ term, and since the LOB also has LOs placed by other market participants, the market makers orders are not always getting executed when a MO arrives. The queue position is modeled through the ρv term, where ρ is called participation rate and it can be interpreted as the probability that an incoming MO will hit our LO standing in the LOB. Ideally, the queue position should also be taken into account, but this would require us to keep track of the queue lengths and the positions in the queue which would make the model more complex in terms of the state space. For the second integral, the logic is similar but since the market maker is buying at the bid, instead of receiving the money, the gain is negative apart from the rebate that the market maker receives from liquidity provision. The sum term in the cash process is related to the MO impulses sent by the market maker. $(S^a_{\tau_n^-} + \eta)\zeta_n^+$ denotes the money paid for sending a buy MO of size $\zeta_n^+ > 0$ plus the exchange fee η paid for liquidity taking, and $(S^b_{\tau_n^+} - \eta)\zeta_n^-$ follows the same logic but for market sell orders with $\zeta_n^- < 0$. The sum term includes the assumption that the MOs sent by the market maker are always small enough so that they do not walk the LOB.

The market makers inventory develops according to

$$
\begin{aligned}
Q_t = q_{t_0} &- \int_{(t_0,t]\times\mathbb{R}_+} \rho v R^a_{r^-}(N_1 + N_7)(dr \times dv) \\
&+ \int_{(t_0,t]\times\mathbb{R}_+} \rho v R^b_{r^-}(N_2 + N_8)(dr \times dv) \\
&+ \sum_{\tau_n \in (t_0,t]} \zeta_n,
\end{aligned}
$$

which is just the initial inventory minus the assets sold at the ask plus the assets bought at the bid and the sum of the inventory gains from hedging.

Now the market makers value function can be defined as

$$V_t(t, x_t) = \sup_{u \in \mathbb{U}} V_t^u(t, x_t^u),$$

where $x_t^u = (B_t^u, Q_t^u, S_t^b, S_t^a, R_t^a, R_t^b)$ is a shorthand notation for the state space and

$$V_t^u(t, x_t^u) = \mathbb{E}\left\{ G(B_T^u, Q_T^u, S_T^b, S_T^a) - \int_{t_0}^{T} H(t, q_z^u)dz - F(S_T^b, S_T^a, R_T^b, R_T^a) \right\}$$

is the market makers performance criteria. The market makers wealth at the end of the trading session is defined as

$$G(B_T^u, Q_T^u, S_T^b, S_T^a) = B_T^u + (S_T^b - \eta)Q_T^+ - (S_T^a + \eta)Q_T^-,$$

where B_T is the cash at the end, $(S_T^b - \eta)Q_T^+$ is the mark to market value of the market makers positive inventory minus the exchange fee η and $(S_T^a + \eta)Q_T^-$ is the mark to market value of the market makers negative inventory plus the exchange fee η. The running inventory penalty is defined as

$$H(t, q_t^u) = -\theta(q_t^u)^2,$$

where $\theta > 0$ is the factor that determines the severity of the penalization. The final term in the value function represents the total cost from the impulses submitted by the market maker i.e. the LOs and MOs sent as well as the cancelled LOs, and is defined as

$$F(S_T^b, S_T^a, R_T^b, R_T^a) =$$

$$\sum_{\tau_n \in (t_0, T]} c^b(r^b_{n-1}, r^b_n, \tau_n, S^a_{\tau_n} - S^b_{\tau_n}) + c^a(r^a_{n-1}, r^a_n, \tau_n, S^a_{\tau_n} - S^b_{\tau_n}) + c^i \mathbb{1}(\zeta_n \neq 0).$$

The $c^a(i, j, t, s)$ and $c^b(i, j, t, s)$ terms define the costs related to switching the regime from i to j, and c^i defines the cost related to the MOs sent by the market maker. The mathematical definitions and more indepth explanations regarding the cost functions are given in the following section.

4.3.2 Cost Functions

The cost functions for switching the LO regimes are defined as follows:

$$c^b(0, 1, t, s) = \alpha \bar{q}^b \rho(s/2 + \epsilon) \min\left\{\frac{T - t}{\bar{q}^b/(\lambda_2 \bar{v}_2 + \lambda_8 \bar{v}_8)}, 1\right\},$$

$$c^a(0, 1, t, s) = \alpha \bar{q}^a \rho(s/2 + \epsilon) \min\left\{\frac{T - t}{\bar{q}^a/(\lambda_1 \bar{v}_1 + \lambda_7 \bar{v}_7)}, 1\right\},$$

$$c^b(1, 0, t, s) = c^a(1, 0, t, s) = 0,$$

where $\bar{v}_i$ is the average volume of order type i, $\bar{q}^b$ and $\bar{q}^a$ are the queue lengths ahead of our market maker in the switching off mode, and α is a discount factor for switching. The cost functions can be interpreted as follows: when removing quotes from the LOB i.e. switching from 1 to 0, there is no cost associated since the market maker just cancels the existing limit orders, and when switching on, the cost resulting from the price time priority mechanism has to be accounted for. $\alpha \bar{q} \rho(s/2 + \epsilon)$ is the cost that results from the fact that the orders ahead of the market makers order have to be executed in order for the market maker to complete the round trip of trades. The minimum term in the cost functions decreases the switching cost closer to the end of the trading session as its preferred by the market maker to liquidate excess inventory passively by quoting top of the book rather than paying the final liquidation cost resulting from a possibly large MO.

The cost for sending MOs is the exchange fee η plus the cost of crossing the spread, which are both incorporated in the market makers cash process, plus an additional slippage cost resulting from the price moving against the market maker before submitting a MO. The slippage cost is defined as follows

$$c^i = \beta \delta \rho \bar{v}_{max},$$

where $\bar{v}_{\max}$ is the maximum average volume of order types 1, 2, 7 and 8, and β is a discount factor for the MO impulses which can be interpreted as the probability that the price moves away one tick before the market makers submitted MO is executed.

As the authors of [27] state, the market makers optimal quoting strategy is quite sensitive to the switching and impulse costs c^b, c^a and c^i and the functions need more research. For the numerical example of the German intraday market, we have chosen the discount factors α and β so that the market maker will always prefer switching LO regimes instead of sending MOs due to the large average spread of the intraday market.

4.3.3 Stochastic Optimal Control Problem

In this section we will look at how the development of the system and the market makers decisions affect the value function during the trading session. Based on these, we show that the market makers value function has to satisfy a Hamilton-Jacobi-Bellman quasi-variational inequality (HJBQVI) which can then be solved numerically to obtain the set of optimal controls. As stated previously, on each time step the market maker can choose to send an impulse that changes the market makers current positioning in the LOB, hedges excess inventory using a MO or both. To find the optimal action the market makers seeks a tuple $(\tilde{r}^b, \tilde{r}^a, \tilde{\zeta}) \in \{0,1\}^2 \times \mathbb{I} \backslash \{(r^a, r^b, 0)\}$ that maximizes the value function, where $\mathbb{I}$ is a signed number depicting the MO volume which has to be a multiple of the traded assets lot size. The optimal impulse is found by shifting the value function by the intervention operator $\mathcal{M}$ which is defined as

$$\mathcal{M}V(t, b, q, s^b, s^a, r^a, r^a) =$$

$$\max_{(\tilde{r}^b, \tilde{r}^a, \tilde{\zeta}) \in \{0,1\}^2 \times \mathbb{I} \backslash \{(r^b, r^a, 0)\}} \big\{ V(t, b - (s^a + \eta)\zeta^+ + (s^b - \eta)\zeta^-, q + \zeta, s^b, s^a, \tilde{r}^b, \tilde{r}^a)$$

$$- c^b(r^b, \tilde{r}^b, t, s^a - s^b) - c^a(r^a, \tilde{r}^a, t, s^a - s^b) - c^i \mathbb{1}(\zeta \neq 0) \big\}.$$

In the case that the market maker does nothing but let the system fluctuate, the value function is shifted by the infinitesimal generator which is defined as

$$\mathcal{L}V(t, b, q, s^b, s^a, r^b, r^a) =$$

$$\int_{\mathbb{R}_+} \sum_{\xi=1}^{\infty} \big(V(t, b_+, q_-, s^b, s^a_+, r^b, r^a) - V(t, b, q, s^b, s^a, r^b, r^a) \big) \lambda_1(t) f_1(v, \xi) dv$$

$$+ \int_{\mathbb{R}_+} \sum_{\xi=1}^{\infty} \big(V(t, b_-, q_+, s^b_-, s^a, r^b, r^a) - V(t, b, q, s^b, s^a, r^b, r^a) \big) \lambda_2(t) f_2(v, \xi) dv$$

$$+ \sum_{\xi=1}^{((s^a - s^b)/\delta) - 1} \big(V(t, b, q, s^b_+, s^a, r^b, r^a) - V(t, b, q, s^b, s^a, r^b, r^a) \big) \lambda_3(t) f_3(\xi)$$

$$+ \sum_{\xi=1}^{((s^a - s^b)/\delta) - 1} \big(V(t, b, q, s^b, s^a_-, r^b, r^a) - V(t, b, q, s^b, s^a, r^b, r^a) \big) \lambda_4(t) f_4(\xi)$$

$$+ \sum_{\xi=1}^{\infty} \big(V(t, b, q, s^b_-, s^a, r^b, r^a) - V(t, b, q, s^b, s^a, r^b, r^a) \big) \lambda_5(t) f_5(\xi)$$

$$+ \sum_{\xi=1}^{\infty} \big(V(t, b, q, s^b, s^a_+, r^b, r^a) - V(t, b, q, s^b, s^a, r^b, r^a) \big) \lambda_6(t) f_6(\xi)$$

$$+ \int_{\mathbb{R}_+} \big(V(t, b_+, q_-, s^b, s^a, r^b, r^a) - V(t, b, q, s^b, s^a, r^b, r^a) \big) \lambda_7(t) f_7(v) dv$$

$$+\int_{\mathbb{R}_+} \big(V(t, b_-, q_+, s^b, s^a, r^b, r^a) - V(t, b, q, s^b, s^a, r^b, r^a)\big)\lambda_8(t) f_8(v) dv,$$

where

$$b_- = b - r^b(s^b - \epsilon)\rho v,$$

$$b_+ = b + r^a(s^a + \epsilon)\rho v,$$

$$q_- = q - r^a \rho v,$$

$$q_+ = q + r^b \rho v,$$

$$s^b_- = s^b - \xi\delta,$$

$$s^b + = s^b + \xi\delta,$$

$$s^a_- = s^a - \xi\delta$$

and

$$s^a_+ = s^a + \xi\delta.$$

Although the expression for the infinitesimal generator might look complicated, it simply records the average change in the value function given that the market maker is posted on both sides of the LOB and is not sending any impulses. For example, the first integral term in the expression records the average change in the value function over the tick and volume distributions of aggressive market buy order. When an aggressive market buy order arrives, the market makers cash increases to $b + r^a(s^a + \epsilon)\rho v$, inventory decreases to $q - r^a \rho v$ and the ask price goes up to $s^a + \xi\delta$. And finally the average change of the value function is scaled by $\lambda_1(t)$ to account for the arrival rate of the given order type.

Now the market makers problem becomes a problem of optimal stopping, i.e. when to issue impulses. It can be shown that the value function is the solution to the following HJBQVI according to [27]

$$\min\big\{-\partial_t V(t, x) - \mathcal{L}V(t, x) + \theta q^2,\ V(t, x) - \mathcal{M}V(t, x)\big\} \tag{5}$$

with a terminal condition

$$V(T, x) = b_T + (s^b_T - \eta)q^+_T - (s^a_T + \eta)q^-_T,$$

but the rigorous proof is omitted due to the fact that we are satisfied with the numerical solution presented in the following sections which allows us to find the optimal quoting strategy.

4.3.4 Solving the Optimal Controls

To solve the HJBQVI of equation (5), the following steps are completed: first we make an educated guess of the general form of V also known as an ansatz, which reduces the number of state variables by two. This is similar to the procedure in, for example, [14]. The ansatz for equation (5) is

$$V(t,x) = b + s^m q + \Phi(t,q,s,r^b,r^a), \tag{6}$$

where $s^m = (s^b + s^a)/2$ and $s = s^a - s^b$. Then after substituting (6) into (5), we find that $\Phi(t,q,s,r^b,r^a)$ satisfies the HJBQVI

$$\min\big\{-\partial_t \Phi(t,x) - \mathcal{L}\Phi(t,x) + \theta q^2,\ \Phi(t,x) - \mathcal{M}\Phi(t,x)\big\}, \tag{7}$$

with terminal condition

$$\Phi(T,x) = -(s/2+\eta)|q|, \tag{8}$$

where x is now a shorthand for the reduced state space $x = (q,s,r^b,r^a)$, $\mathcal{L}\Phi(t,x)$ is defined as

$$\mathcal{L}\Phi(t,q,s,r^b,r^a) =$$

$$\int_{\mathbb{R}_+} \sum_{\xi=1}^{\infty} \big(\Phi(t,q_-,s_+,r^b,r^a) - \Phi(t,q,s,r^b,r^a) + \xi\delta q_-/2 + r^a \rho v(s/2+\epsilon)\big)\lambda_1(t) f_1(v,\xi)dv$$

$$+ \int_{\mathbb{R}_+} \sum_{\xi=1}^{\infty} \big(\Phi(t,q_+,s_+,r^b,r^a) - \Phi(t,q,s,r^b,r^a) - \xi\delta q_+/2 + r^b \rho v(s/2+\epsilon)\big)\lambda_2(t) f_2(v,\xi)dv$$

$$+ \sum_{\xi=1}^{(s/\delta)-1} \big(\Phi(t,q,s_-,r^b,r^a) - \Phi(t,q,s,r^b,r^a) + \xi\delta q/2\big)\lambda_3(t) f_3(\xi)$$

$$+ \sum_{\xi=1}^{(s/\delta)-1} \big(\Phi(t,q,s_-,r^b,r^a) - \Phi(t,q,s,r^b,r^a) - \xi\delta q/2\big)\lambda_4(t) f_4(\xi)$$

$$+ \sum_{\xi=1}^{\infty} \big(\Phi(t,q,s_+,r^b,r^a) - \Phi(t,q,s,r^b,r^a) - \xi\delta q/2\big)\lambda_5(t) f_5(\xi)$$

$$+ \sum_{\xi=1}^{\infty} \big(\Phi(t,q,s_+,r^b,r^a) - \Phi(t,q,s,r^b,r^a) + \xi\delta q/2\big)\lambda_6(t) f_6(\xi)$$

$$+ \int_{\mathbb{R}_+} \big(\Phi(t,q_-,s,r^b,r^a) - \Phi(t,q,s,r^b,r^a) + r^a \rho v(s/2+\epsilon)\big)\lambda_7(t) f_7(v)dv$$

$$+ \int_{\mathbb{R}_+} \big(\Phi(t,q_+,s,r^b,r^a) - \Phi(t,q,s,r^b,r^a) + r^b \rho v(s/2+\epsilon)\big)\lambda_8(t) f_8(v)dv,$$

where $s_- = s - \xi\delta$ and $s_+ = s + \xi\delta$, and $\mathcal{M}\Phi(t,x)$ is defined as

$$\mathcal{M}V(t,q,s,r^b,r^a) = \max_{(\tilde{r}^b,\tilde{r}^a,\tilde{\zeta}) \in \{0,1\}^2 \times \mathbb{I} \setminus \{(r^b,r^a,0)\}} \big\{\Phi(t,q+\zeta,s,\tilde{r}^b,\tilde{r}^a)$$

$$-c^b(r^b,\tilde{r}^b,t,s)-c^a(r^a,\tilde{r}^a,t,s)-c^i\mathbb{1}(\zeta\neq 0)-(s/2+\eta)|\zeta|\big\}.$$

Both of the expression for $\mathcal{M}\Phi(t,x)$ and $\mathcal{L}\Phi(t,x)$ can be obtained by substituting the ansatz for V into their respective definitions.

In order to solve equation (7), we will apply a numerical method called a penalty scheme, for more details see [27]. The penalty scheme states that with a sufficiently small $\gamma > 0$, equation (7) can be approximated by the following representation

$$\min\big\{-\partial_t\Phi-\mathcal{L}\Phi+\theta q^2,\ \Phi-\mathcal{M}\Phi-\gamma(\partial_t\Phi-\mathcal{L}\Phi+\theta q^2)\big\}=0,$$

which is equivalent to

$$\min\big\{-\partial_t\Phi-\mathcal{L}\Phi+\theta q^2,\ (\Phi-\mathcal{M}\Phi)/\gamma-(\partial_t\Phi-\mathcal{L}\Phi+\theta q^2)\big\}=0,$$

and thus also to

$$-\partial_t\Phi-\mathcal{L}\Phi+\theta q^2+\min\big\{0,(\Phi-\mathcal{M}\Phi)/\gamma\big\}=0.$$

Finally, we arrive at

$$-\partial_t\Phi-\mathcal{L}\Phi+\theta q^2-(\mathcal{M}\Phi-\Phi)^+/\gamma=0,$$

where $\Phi=\Phi(t,x)$ for shorter notation and $(a)^+=\max\{a,0\}$. By replacing the time derivative $\partial_t\Phi$ with a discretized version $\partial_t\Phi\approx\frac{\Phi(t_n)-\Phi(t_{t-1})}{dt}$ we get

$$-\frac{\Phi(t_n)-\Phi(t_{n-1})}{dt}-\mathcal{L}\Phi(t_n)+\theta q^2-(\mathcal{M}\Phi(t_n)-\Phi(t_n))^+/\gamma=0,$$

from which we can solve the feedback formula

$$\Phi(t_{n-1})=\Phi(t_n)+dt(\mathcal{L}\Phi(t_n)-\theta q^2)-\frac{dt}{\gamma}(\mathcal{M}\Phi(t_n)-\Phi(t_n))^+,$$

where the terminal condition is the same as before

$$\Phi(T,x)=-(s/2+\epsilon)|q|.$$

In order to solve the values of Φ in practice, we will define a finite grid of time, spread, inventory and bid and ask regimes. The integrals inside $\mathcal{L}\Phi$ can be estimated using Simpsons rule for numerical integration on a bounded interval $(0,\chi_i(-\log(0.01)^{1/\kappa_i})$ and the values of $\mathcal{M}\Phi$ can be found by exhaustive search over the optimization space $\{0,1\}^2\times\mathbb{I}$. The values of $\Phi(q_-)$, $\Phi(q_+)$, $\Phi(s_-)$ and $\Phi(q_+)$ that do not fall on the grid can be estimated by linear interpolation.

Having solved Φ, we can now find the optimal controls from

$$\tau_k=\min\big\{t_n>\tau_{k-1}\mid\Phi(t_n,Q_{t_n},S_{t_n},R^b_{t_n},R^a_{t_n})\leq\mathcal{M}\Phi(t_n,Q_{t_n},S_{t_n},R^b_{t_n},R^a_{t_n})\big\}\tag{9}$$

where the rationale is that we just select the next stopping time larger than τ_{k-1} for which the value of Φ is larger after an impulse than it was before i.e. sending an impulse improves the expected total wealth at the end.

4.3.5 Action Thresholds

Although equation (9) returns the stopping times and the optimal actions at those times as a function of the full state space, from the practical point of view it would be preferred that we have a simple lookup table which gives us continuous limits for LO additions, cancellations and sending MOs as a function of time and spread. Equations (10)-(17) define how such thresholds can be obtained using the previously solved values of Φ:

$$q^b_{\text{off}}(t,s) = \inf\left\{q \in \mathbb{R} \,\middle|\, \Phi(t,q,s,0,1) - c^b(1,0,t,s) \geq \Phi(t,q,s,1,1)\right\}, \tag{10}$$

$$q^b_{\text{imp}}(t,s) = \min\Bigg\{ \tag{11}$$

$$\inf\left\{q \geq q^b_{\text{off}}(t,s) \,\middle|\, \Phi(t,q+\zeta,s,0,1) + \zeta\left(\frac{s}{2}+\eta\right) - c^i \geq \Phi(t,q,s,0,1),\ \exists\,\zeta < 0\right\},$$

$$\inf\left\{q \in \mathbb{R} \,\middle|\, \Phi(t,q+\zeta,s,1,1) + \zeta\left(\frac{s}{2}+\eta\right) - c^i \geq \Phi(t,q,s,1,1),\ \exists\,\zeta < 0\right\}$$

$$\Bigg\},$$

$$q^b_{\text{action}}(t,s) = \min\left\{q^b_{\text{off}}(t,s),\ q^b_{\text{imp}}(t,s)\right\}, \tag{12}$$

$$q^b_{\text{on}}(t,s) = \sup\left\{q \leq q^b_{\text{off}}(t,s) \,\middle|\, \Phi(t,q,s,1,1) - c^b(0,1,t,s) \geq \Phi(t,q,s,0,1)\right\}. \tag{13}$$

Equations (10)-(13) define the thresholds for the bid-side, and they can be interpreted as follows: when the market makers inventory exceeds q^b_{off} at time t, the market maker should cancel existing buy LOs and stop buying until inventory goes below q^b_{on}, after which the market maker can add buy LOs to the LOB and continue to be posted on the bid-side. In the case that $q^b_{\text{off}} > q^b_{\text{imp}}$ at time t, the market maker should prefer to reduce positive inventory using sell MOs while still being active on the bid-side. The thresholds are similarly defined for the ask side in equations (14)-(17).

$$q^a_{\text{off}}(t,s) = \sup\left\{q \in \mathbb{R} \mid \Phi(t,q,s,1,0) - c^a(1,0,t,s) \geq \Phi(t,q,s,1,1)\right\}, \tag{14}$$

$$q^a_{\text{imp}}(t,s) = \max\Bigg\{ \tag{15}$$

$$\sup\left\{q \leq q^a_{\text{off}}(t,s) \,\middle|\, \Phi(t,q+\zeta,s,1,0) - \zeta\left(\frac{s}{2}+\eta\right) - c^i \geq \Phi(t,q,s,1,0),\ \exists\,\zeta > 0\right\},$$

$$\sup\left\{q \in \mathbb{R} \,\middle|\, \Phi(t,q+\zeta,s,1,1) - \zeta\left(\frac{s}{2}+\eta\right) - c^i \geq \Phi(t,q,s,1,1),\ \exists\,\zeta > 0\right\}$$

$$\Bigg\},$$

$$q^a_{\text{action}}(t,s) = \max\left\{q^a_{\text{off}}(t,s),\ q^a_{\text{imp}}(t,s)\right\}, \tag{16}$$

$$q^a_{\text{on}}(t,s) = \\ \inf\left\{q \geq q^a_{\text{off}}(t,s) \;\middle|\; \Phi(t,q,s,1,1) - c^b(0,1,t,s) \geq \Phi(t,q,s,1,0)\right\}. \tag{17}$$

The logic for the ask side is similar to the bid side but reversed, meaning that when the market makers inventory goes below q^a_{off}, they should cancel existing sell orders and stop selling until the inventory goes above q^a_{on}, and if $q^a_{\text{off}} < q^a_{\text{imp}}$, the market maker should reduce negative inventory using buy MOs instead of passive buying, while staying posted on the ask side.

5 Numerical Example

In this section we will study the behavior of the market making model in an ideal scenario with symmetrical order arrival rates in the German intraday market. The model was solved from the end of IDA 2 until the start of IDA 3. The reason for this choice was that the arrival rates begin to increase during this period while the liquidity at the top of the book still remains at a reasonable level compared to the end of the trading session as seen in Figures 3.4 and 3.5. Table 2 presents the arrival rate parameters for all the order types.

Arrival rate	ν	ψ	ω
λ_1, λ_2	0.00155	0.000024	0.00018
λ_3, λ_4	0.05905	0.0445608	0.00007
λ_5, λ_6	0.01491	0.0033175	0.00009
λ_7, λ_8	0.01067	0.0001691	0.00015

Table 2: Arrival rate parameters for different order types from Figure 4.1.

Table 3 shows exchange, model and discretization parameters needed for solving the model. The exchange parameters are from EPEX SPOT where the tick size is 1 cent and the exchange does not pay rebate for liquidity provision. Choice of the model parameters is two fold: $\bar{q}^b$, $\bar{q}^a$ and ρ were chosen based on the empirical analysis, and θ, α and β were chosen in order to get a set of thresholds where LO cancellation is preferred over MOs. θ does not affect the shape of the thresholds in other ways but widening them in case the parameter is decreased.

Exchange Parameter	Value	Model Parameter	Value	Discretization Parameter	Value
δ	0.01	θ	1e-5	T	39 600s
ϵ	0	ρ	0.04	dt	1s
η	0.008	$\bar{q}^b$	5 MW	h_q	0.1 MW
		$\bar{q}^a$	5 MW	N_q	101
		α	1	N_s	2
		β	0	h_l	0.2 MW
				h_M	0.1 MW
				γ	1

Table 3: Model, exchange and discretization parameters.

Figure 5.1 shows the inventory thresholds for the market makers actions as a function of time and inventory for a spread equal to one tick and Figure 5.2 presents the same inventory thresholds for the final 1600 seconds of the trading session. As can be seen, the inventory limits keep widening as the trading session progresses, this is expected given that the order arrivals also increase which means that it is easier for the market maker to get rid of excess inventory. The impulse threshold is always above (below) the bid-off (ask-off) threshold, so the market maker will always prefer

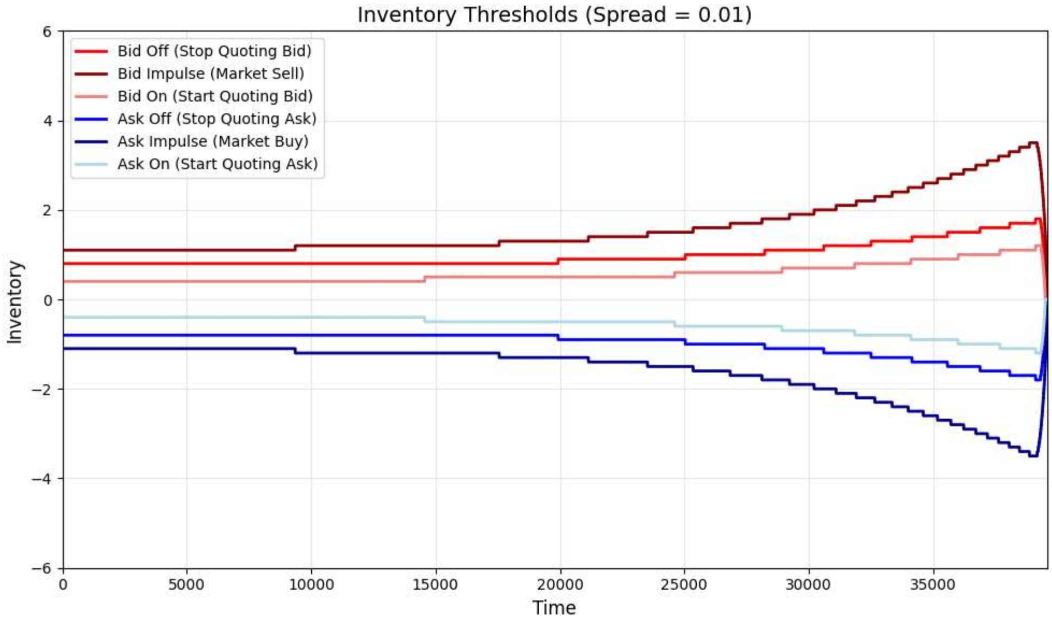

Figure 5.1: Optimal LO and MO thresholds for the whole trading period.

passively selling (buying) instead of sending MOs. The only case in which MOs might actually be used, according to these thresholds, is towards the end of the trading session if the market makers inventory goes beyond either the bid-off or ask-off thresholds, but the passive LOs on the other side are not executed before the impulse threshold is crossed. All of the thresholds keep increasing as the trading session advances which

Figure 5.2: Optimal LO and MO thresholds for the final 1600s seconds of the trading session.

matches the increasing liquidity curve of the intraday market, and just a few minutes before the end of the trading session all of the thresholds cut back to zero since the market maker wants to get rid of the inventory passively.

Figure 5.3 shows the sizes of the MOs sent by the market maker for the thresholds in Figures 5.1 and 5.2. The MOs stay at 0.1 MW for the negative inventory side and at -0.1 MW for the positive inventory side for a long time, until around 30 000 seconds when the size starts to increase/decrease along with the impulse thresholds. The size of the MOs stays relatively small for the whole trading session and in particular, the size is never large enough that the inventory would become zero.

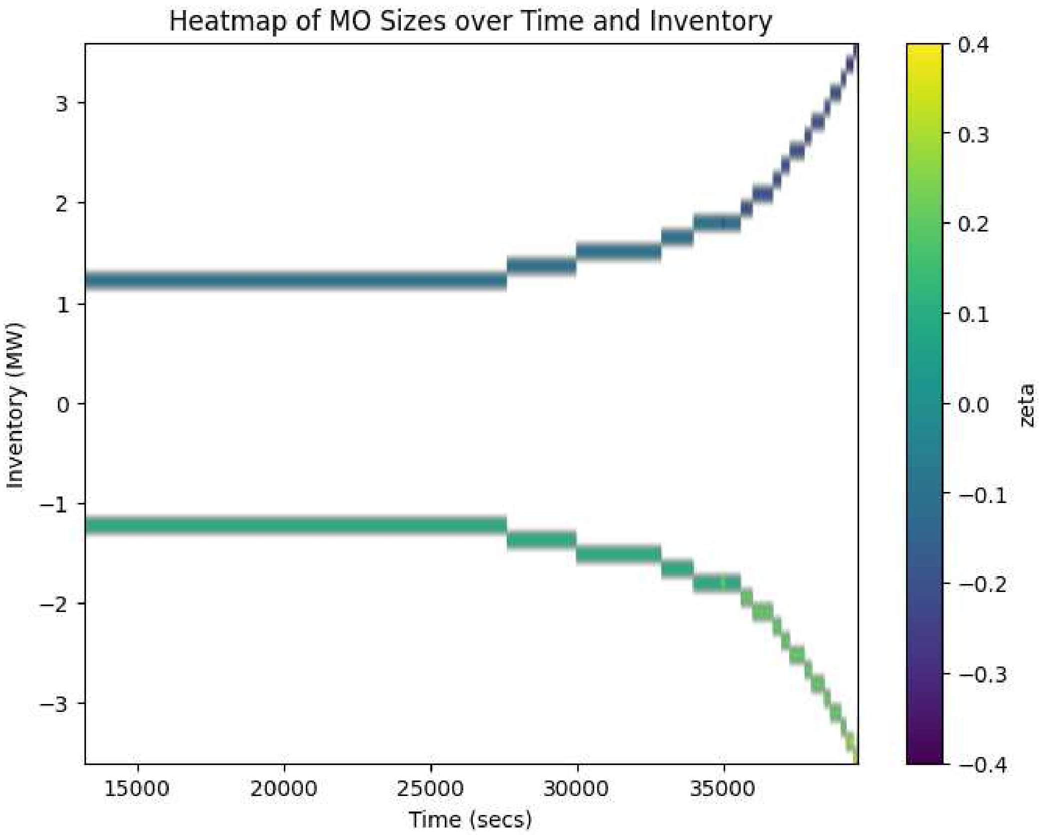

Figure 5.3: Heatmap of MO order sizes sent by the market maker vs. time and inventory.

A few practical considerations regarding the model are in order: firstly the model is quite unstable numerically. In the example case we only had two grid points for the spread, but if we increase the number, the values of the value function quickly grow to infinity, and as a result of this, we are unable to obtain the inventory thresholds. The numerical instability is most likely caused by the large differences in the arrival rates, so solving the model for a shorter time period or with a smaller time step could possibly fix this. Secondly, whether we should always quote top of the book or perform some kind of spread based optimization in an unliquid market is an important question and is a matter of how much value we give to accurately modeling the microstructure.

A large part of the modeling assumptions made in the original paper [27] assume a highly liquid market with strong queues which do not hold true for the intraday continuous market. Finally, the model parameters α, β are quite arbitrary and difficult to estimate in practice, while they also have a large effect on the inventory thresholds. Because of this, the cost functions should probably be defined in a simpler way that is not time dependent.

6 Conclusions

Due to the significant growth in renewables integration, the importance of the intraday continuous power market has been continuously increasing in the recent years. Understanding the determinants of liquidity and price formation is a key topic for market participants in order to develop optimal trading strategies. Increasing the liquidity of the intraday continuous market is beneficial from the perspective of the overall market quality. In this thesis we developed a market making model for the German intraday power market and conducted empirical analysis of its microstructure.

Our empirical analysis replicates key findings from the literature on liquidity determinants and extends them by examining the effects of intraday auctions on the market quality and by approximating the instantaneous market impact of aggressive orders. Building on these insights, we introduced a stochastic impulse-control market-making framework that addresses several microstructural limitations of earlier models in the Avellaneda–Stoikov framework. By employing a weakly consistent order book model and incorporating empirically motivated features, most notably the increasing arrival rates that reflect the liquidity curve of the intraday market, we obtain a model that better captures the trading environment faced by actual liquidity providers. Numerical examples illustrate how the optimal quoting strategy adapts to these microstructural features. A limitation of the model is its numerical instability that arises for long time horizons or large time steps, and requires careful parameter choices in order to obtain tangible results.

Future work could focus on deriving analytical or semi-analytical approximations for the market making problem, ideally within a framework where order arrivals follow Hawkes processes. Such developments would allow short-term predictions of order flow and enable a more flexible incorporation of future price signals than is currently possible with a purely numerical solution.

References

[1] *Share of energy from renewable sources*. Eurostat. URL: https://ec.europa.eu/eurostat/databrowser/view/NRG_IND_REN/default/table?lang=en (visited on Sept. 17, 2025).

[2] *Annual Trading Results of 2024*. EPEX SPOT. URL: https://www.epexspot.com/sites/default/files/download_center_files/2025-01-28_EPEX%20SPOT_Annual%20Power%20Trading%20Results%202024_finaldraft_0.pdf (visited on Sept. 11, 2025).

[3] R. Weron. *Complex Electricity Markets*. John Wiley & Sons, Ltd, 2006. Chap. 1, pp. 1–24.

[4] *EU Market Integration*. EnExGroup. URL: https://www.enexgroup.gr/web/guest/eu-market-integration (visited on Sept. 11, 2025).

[5] P. Pinson. *Renewables in Electricity Markets, Module 2, Section 2.1*. DTU. URL: https://pierrepinson.com/31761/Slides/31761lecture2p1.pdf (visited on Sept. 12, 2025).

[6] *EUPHEMIA: Description and functioning*. EPEX SPOT. URL: https://www.epexspot.com/sites/default/files/2020-02/Euphemia_Description%20and%20functioning_1812.pdf (visited on Sept. 12, 2025).

[7] *Intraday Auctions on SIDC*. ENTSO-E. URL: https://www.entsoe.eu/network_codes/cacm/implementation/ida/ (visited on Sept. 13, 2025).

[8] *Cross-Border Intraday: Questions Answers*. ENTSO-E. URL: https://eepublicdownloads.entsoe.eu/clean-documents/Network%20codes%20documents/Implementation/cacm/xbid/XBID%20Q-A.pdf (visited on Sept. 13, 2025).

[9] *PICASSO*. ENTSO-E. URL: https://www.entsoe.eu/network_codes/eb/picasso/ (visited on Sept. 18, 2025).

[10] *MARI*. ENTSO-E. URL: https://www.entsoe.eu/network_codes/eb/mari/ (visited on Dec. 23, 2025).

[11] I. Roșu. "A dynamic model of the limit order book". In: *Review of Financial Studies* 22.11 (Nov. 2009), pp. 4601–4641.

[12] M. D. Gould et al. "Limit order books". In: *Quantitative Finance* 13.11 (2013), pp. 1709–1742.

[13] L. R. Glosten and P. R. Milgrom. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders". In: *Journal of Financial Economics* 14.1 (1985), pp. 71–100.

[14] M. Avellaneda and S. Stoikov. "High Frequency Trading in a Limit Order Book". In: *Quantitative Finance* 8 (Apr. 2008), pp. 217–224. DOI: 10.1080/14697680701381228.

[15] T. Ho and H. R. Stoll. "Optimal dealer pricing under transactions and return uncertainty". In: *Journal of Financial Economics* 9.1 (1981), pp. 47–73.

[16] P. Fodra and H. Pham. "High frequency trading and asymptotics for small risk aversion in a Markov renewal model". In: (2015). arXiv: 1310.1756 [q-fin.TR]. URL: https://arxiv.org/abs/1310.1756.

[17] P. Fodra and M. Labadie. "High-frequency market-making with inventory constraints and directional bets". In: (2012). arXiv: 1206.4810 [q-fin.TR]. URL: https://arxiv.org/abs/1206.4810.

[18] P. Fodra and M. Labadie. *High-frequency market-making for multi-dimensional Markov processes*. 2013. arXiv: 1303.7177 [q-fin.TR]. URL: https://arxiv.org/abs/1303.7177.

[19] F. Pomponio and F. Abergel. "Trade-throughs: Empirical Facts and Application to Lead-lag Measures". In: *Econophysics of Order-driven Markets: Proceedings of Econophys-Kolkata V*. Springer Milan, 2011, pp. 3–16.

[20] J.-P. Bouchaud et al. *Trades, Quotes and Prices: Financial Markets Under the Microscope*. Cambridge University Press, 2018.

[21] S. Hagemann and C. Weber. "An Empirical Analysis of Liquidity and its Determinants in The German Intraday Market for Electricity". In: *Energy Economics* (2013).

[22] C. Balardy. "An Empirical Analysis of the Bid-ask Spread in the Continuous Intraday Trading of the German Power Market". In: *The Energy Journal* 43.3 (2022), pp. 229–256. DOI: 10.5547/01956574.43.3.cbal.

[23] H. Demsetz. "The Cost of Transacting". In: *The Quarterly Journal of Economics* 82.1 (1968), pp. 33–53. DOI: https://doi.org/10.2307/1882244.

[24] N. Graf von Luckner and R. Kiesel. "Modeling Market Order Arrivals on the German Intraday Electricity Market with the Hawkes Process". In: *Journal of Risk and Financial Management* 14.4 (2021). DOI: 10.3390/jrfm14040161.

[25] N. Torre. *BARRA Market Impact Model Handbook*. BARRA Inc, 1997.

[26] C. Kath and F. Ziel. *Optimal Order Execution in Intraday Markets: Minimizing Costs in Trade Trajectories*. 2020. arXiv: 2009.07892. URL: https://arxiv.org/abs/2009.07892.

[27] B. Law and F. Viens. *Market Making under a Weakly Consistent Limit Order Book Model*. 2020. URL: https://arxiv.org/abs/1903.07222.

[28] K. Dayri and M. Rosenbaum. "Large tick assets: implicit spread and optimal tick size". In: *Market Microstructure and Liquidity* 1.01 (2015), p. 1550003.