

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
ECOLE POLYTECHNIQUE DE LOUVAIN

CENTER FOR OPERATIONS RESEARCH AND ECONOMETRICS



Enhancing the Value of Flexible Generation Capacity through Continuous Intraday Markets and Scarcity Pricing

Gilles Bertrand

Doctoral Thesis

Supervisor:

Anthony Papavasiliou
(UCLouvain, Belgium)

Jury:

Philippe Chevalier (UCLouvain, Belgium)
Pierre-Antoine Absil (UCLouvain, Belgium)
Stein-Erik Fleten (NTNU, Norway)
Mads Trolle (Centrica Energy Trading, Denmark)

President:

Philippe Chevalier (UCLouvain, Belgium)

PhD Organization

Gilles Bertrand

Université catholique de Louvain
École Polytechnique de Louvain
Center of Operations Research and Econometrics

Thesis Supervisor

Anthony Papavasiliou

Associate Professor, Université catholique de Louvain
École Polytechnique de Louvain
Center of Operations Research and Econometrics

Supervisory Committee

Yves Smeers

Emeritus professor, Université catholique de Louvain
École Polytechnique de Louvain
Center of Operations Research and Econometrics

Bertrand Cornélusse

Assistant Professor, Université de Liège
Institut Montefiore

Abstract

The increasing integration of intermittent renewable production requires more flexible assets in power systems. There are two potential paths in order to increase the remuneration of these flexible assets. The first one is to improve trading strategies in existing markets. The second one is to introduce new market mechanisms. This thesis studies both aspects by developing trading strategies for the Continuous Intraday Market, and by analyzing different options for implementing a “scarcity pricing mechanism” in the European electricity market design.

The contributions of the dissertation are organized in three chapters. Chapter 2 presents a method for trading the production of a storage unit in the Continuous Intraday Market. We model this problem in the Markov Decision Process framework. We present an approach based on Policy Function Approximation for tackling the problem. We provide relevant parameters for defining our policy, and demonstrate the effectiveness of our approach by comparing it to a method commonly employed in the industry on real historical data. In chapter 3, we are interested in the problem of a renewable unit covering its position in the Continuous Intraday Market. As a starting point for tackling this problem, we characterize an optimal policy for trading a fixed quantity in a simplified version of the Continuous Intraday Market. We use this analytical solution as a basis for developing a Value Function Approximation algorithm and an alternative Stochastic Dual Dynamic Programming algorithm that can trade under a more realistic set of assumptions. Chapter 4 proposes a methodology for analysing different market design options for implementing scarcity pricing in the European markets. The methodology relies on analytical insights that can be derived under an assumption of price-taking behavior. These insights are validated by a simulation model which represents the European balancing market as a Markov Decision Process. Our results highlight the benefit of introducing a market in the European market design for trading balancing capacity in real time.

Acknowledgments

First and foremost I am extremely grateful to my supervisor, Anthony. He has all the qualities one can hope from a thesis supervisor. His passion about research inspired me through the thesis and his technical expertise has been invaluable. I am very thankful for the freedom he gave me to explore the topics that interested me the most. He has also been a great source of inspiration for better presenting and structuring my ideas. I am certainly not the same as when I first met him as a master student and he played a big role in that change. For all of that, I will be forever grateful.

I would like to express my gratitude to my jury members, Pr. Philippe Chevalier, Pr. Pierre-Antoine Absil, Pr. Stein-Erik Fleten and Dr. Mads Lund Trolle. Their valuable inputs have considerably improved the quality of the manuscript.

I would like to thank all the members of CORE. Especially, Pr. Yves Smeers for his helpful advices through the thesis. Pr. François Glineur and Pr. Isabelle Thomas for their valuable comments during the preparation of my FRIA proposal. All the participants of the Reinforcement Learning seminars with who it was interesting to dig deeper in the topic. Catherine without who the CORE would not be the CORE. She solves all the problems before they happen.

Thanks to Anthony, I had the chance to confront my ideas with the reality of the industry. I would like to extend my gratitude to Dr. Gauthier de Maere, Dr. Guillaume Erbst, Pr. Damien Ernst, Pr. Bertrand Cornélusse, and Ioannis for the interesting feedback about my work. I was also fortunate to realize an internship at Centrica. I would like to thank all the members of the automated trading department for their warm welcome. A special thanks to Mads for supervising me and giving me helpful insights about the points that should be considered while trading in the real world.

I am also grateful to the members of the energy research group during my stay, Ignacio, Yuting, Ilyès, Céline, Quentin, Loic, Daniel, Jacques, Jehum, and Nicolas. It was nice to spend time with you. A special mention for Yuting with who I attended countless conferences. Thanks for the interesting discussions and for always being in a good mood. We started almost at the same time with Ilyès and we finally made it to the end. It was nice to share the the Ph.D. process with you. I was lucky to have Jacques as my office mate. It has been a pleasure to discuss scarcity pricing and bicycle races with you.

I would like to thank my CREG colleagues for the interesting discussions that helped me improve the manuscript. Especially, Alain for the passionate discussions about all types of market designs, Marijn for his deep understanding of the balancing markets, and Andreas for raising the importance of risk aversion in investment decisions.

I am grateful to my family for their great support during my thesis. My wife Sarah for bearing with me when I was obsessed with a problem I cannot solve. My son for motivating me for a new day of work while he smiles at me in the morning. My mother for the support during all these years and for doing all she can to give me the best chances to do what I want in my life. My grandfather for making me enjoy mathematics and my grandmother for teaching me the importance to have a work ethic.

Finally, the financial support of some institutions has been essential. The research leading to this dissertation has been funded by the Belgian National Science Foundation (FSR-FNRS) through a FRIA grant, by the ENGIE Chair on Energy Economics and Energy Risk Management, and by an Electrabel grant on “Modeling the Value of Flexibility at Sub-Hourly Operating Time Scales”.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Flexibility	3
1.3	Overview of Electricity Markets	4
1.3.1	Day-ahead Auction	5
1.3.2	Intraday Auction	5
1.3.3	Continuous Intraday Market	5
1.3.4	Balancing Markets	6
1.3.5	Imbalance Settlement	7
1.3.6	Scarcity Pricing based on an Operating Reserve Demand Curve	8
1.4	Reinforcement Learning	9
1.4.1	Markov Decision Processes	10
1.4.2	Methods for Solving MDPs	12
1.4.3	Approximation Methods	14
1.5	Contributions	16
1.5.1	Chapter 2	16
1.5.2	Chapter 3	16
1.5.3	Chapter 4	17
2	Optimizing Trading Strategies in the Continuous Intraday Market for a Storage Unit using Reinforcement Learning	19
2.1	Introduction	19
2.1.1	Literature Review	19
2.1.2	Contributions	21
2.1.3	Chapter Organization	21
2.2	Continuous Intraday Market Simulation	21
2.3	Modelling the Intraday Trading Problem Using the MDP Framework	24
2.3.1	State Variables	24
2.3.2	Action Variables	25
2.3.3	Reward	25
2.3.4	State Transition Function	25
2.4	Threshold Policy	25
2.5	Factors Driving the Optimal Threshold	27
2.5.1	Delivery Time	27
2.5.2	Intraday Auction Curve	28
2.5.3	Quantity Already Traded	30

2.5.4	Remaining Time before Market Closure	30
2.5.5	Relative Value of Observable Bids	31
2.5.6	Preventing Imbalances	33
2.5.7	Adapting with Respect to Round-Trip Efficiency	33
2.6	Case Study	34
2.6.1	Rolling Intrinsic Method	34
2.6.2	Learning Process	35
2.6.3	Out-of-sample Testing	36
2.6.4	Profitability of a storage unit trading in the CIM	41
2.7	Conclusions and Perspectives	43
2.8	Appendix	43
2.8.1	Appendix A: Quantity that we exchange compared to the total ex- changed quantity	43
2.8.2	Appendix B: Characterization of Optimal Trading Policy in a De- terministic Setting	44
2.8.3	Appendix C: Rolling Intrinsic Model	46
2.8.4	Appendix D: Order book simulator	48
2.8.5	Determination of Regimes	52
3	Optimal Trading of a Fixed Quantity of Power in an Illiquid Continuous Intraday Market	55
3.1	Continuous Intraday Market Model	56
3.2	Analytical Solutions	58
3.2.1	Assumptions	58
3.2.2	Modelling the Problem as an MDP	58
3.2.3	Optimal Trading Policy	59
3.2.4	Insights from the Analytical Solution	66
3.3	Algorithmic Approaches	68
3.3.1	Value Function Parametrization	68
3.3.2	Stochastic Dual Dynamic Programming	69
3.4	Case Study	69
3.5	Conclusions and Perspectives	71
3.6	Appendix A: Computation of the Value Function	73
3.6.1	Proof of Proposition 1	73
3.6.2	Proof of Proposition 2	73
3.6.3	Proof of Proposition 3	76
3.7	Appendix B: Comparison of the Value Functions	76
3.7.1	Proof of Proposition 4	76
3.7.2	Proof of Proposition 5	79
3.7.3	Proof of Proposition 6	80
4	Market Design Options for Scarcity Pricing in European Balancing Mar- kets	83
4.1	Introduction	83
4.1.1	Comparison between European and US markets	83
4.1.2	Motivation	84
4.1.3	Existing Modeling Frameworks	84
4.1.4	Contributions and Structure	85

4.2	Modeling the Balancing Market in the MDP Framework	86
4.2.1	Problem Description	86
4.2.2	MDP Modelling	87
4.3	Considered Designs	88
4.3.1	Design 1	88
4.3.2	Design 2	88
4.3.3	Design 3	89
4.3.4	Design 4	89
4.3.5	Design 5	90
4.3.6	Design 6	90
4.4	Analysis under Perfect Competition Assumption	91
4.4.1	Generalities	91
4.4.2	Design 1	92
4.4.3	Design 2	97
4.4.4	Design 3	98
4.4.5	Design 4	99
4.4.6	Design 5	100
4.4.7	Design 6	101
4.5	Case study	102
4.5.1	Validation of Analytical Results	102
4.5.2	Back-Propagation	104
4.5.3	Relaxing the Perfect Competition Assumption	105
4.5.4	Other Factors Affecting Balancing Capacity Prices	107
4.6	Conclusion	107
4.7	Appendix A: Computation of the Analytical Solution in the Perfect Com- petition Case	108
4.7.1	Design 2	108
4.7.2	Design 3	110
4.7.3	Design 4	116
4.7.4	Design 5	121
4.7.5	Design 6	127
5	Conclusions and Perspectives	133
5.1	A Summary of Conclusions	133
5.1.1	Case Study on a storage unit trading in the CIM	133
5.1.2	Case Study on the trading of a fixed quantity in the CIM	134
5.1.3	Case Study on the market design of a scarcity pricing mechanism in Belgium	134
5.2	Future Areas of Research	134

List of Figures

1.1	Annual profitability of a CCGT unit of 400 MW in Belgium with an efficiency of 50% and a yearly fixed cost of 7.5 million euros. The production is offered in the forward market, as well as in short-term markets [CREb].	2
1.2	Merit order curve with low renewable infeed (left) and high renewable infeed (right) [PS].	2
1.3	The sequence of operations in a typical central European short-term electricity market.	5
1.4	Illustration of a merit order curve.	7
1.5	Illustration of scarcity pricing [CP21].	9
1.6	Description of the interaction between the agent and the environment in the Reinforcement Learning paradigm ¹ .	11
2.1	Comparison between the best price for selling power that can be achieved when accepting bids versus when placing bids.	22
2.2	Threshold policy for the hydro problem if we consider four possible actions: sell 0, 10, 20 or 30 MWh. The bell curve indicates the probability density function of the sell threshold. The two purple segments and the two red segments of the bell curve indicate the probability of each of the four actions. The green decreasing function corresponds to the buy bids that are available in the order book.	27
2.3	The delivery time of an order impacts its threshold: buying power at 30 €/MWh is not worthwhile in hour 6, but it is worthwhile in hour 17.	28
2.4	Buy regime (left), and sell regime (right) based on the intraday auction price.	29
2.5	Continuous intraday market price for two different days. The curves correspond to different average values, therefore different thresholds need to be applied for an effective threshold strategy.	29
2.6	Illustration of the probability reallocation that relies on the auxiliary Gaussian distribution, as described in section 2.5.5.	32
2.7	Evolution of profit as a function of iterations of the REINFORCE algorithm.	36
2.8	Evolution of α_4^s (left) and the average profit (right) for 6 realizations of the REINFORCE algorithm.	36
2.9	Distribution of the difference between the profit of the threshold policy and the rolling intrinsic policy	38
2.10	Bid acceptance patterns for 1 day of trading for the rolling intrinsic (left) and threshold method (right).	38
2.11	Cumulative payoff evolution for one day of trading for the rolling intrinsic method for various trading frequencies.	40

2.12	Cumulative payoff evolution for one day of trading for the full day (left) and a zoom in on the end of the day (right).	40
2.13	Distribution of the difference between the profit of the threshold policy and the rolling intrinsic policy in the case with round-trip efficiency losses. . . .	41
2.14	Distribution of the difference between the profit of the rolling intrinsic policy and the profit of the threshold policy without parameters $\alpha_3^{s/b}$ and $\alpha_4^{s/b}$. . .	42
2.15	Ratio between our traded quantity and the total exchanged quantity for 10 randomly selected days.	44
2.16	Price in the CIM and in the intraday auction for the first week of 2015. . .	52
3.1	Example of an order book.	56
3.2	Illustration of the different components of the order book.	57
3.3	Estimation of r (left), and expected bid-ask-spread (right).	58
3.4	Analytical solution: quantity that remains to be traded, s_t , with an initial quantity of 25MWh (left) and 100MWh (right).	68
3.5	comparison of value functions for the analytical solution, VFA and SDDP (left). Error in the value function of ADP and SDDP compared to the analytical solution (right).	70
3.6	Quantity that still needs to be traded, s_t , with an initial quantity of 50MWh (left) and 100MWh (right).	70
3.7	Comparison of the error percentage in the value function with respect to the run-time for VFA and SDDP.	71
4.1	Sequence of events faced by an agent in the balancing markets.	87
4.2	The evolution of the balancing capacity price in the simulation of section 4.5.2.	105
4.3	The evolution of the balancing capacity price in the simulation of section 4.5.3 under a setting of imperfect competition.	106

List of Tables

2.1	Percentage of offers that are observed as a function of frequency of accessing the market data.	35
2.2	Profit mean [€/day]	37
2.3	Financial and technical informations on storage units.	43
2.4	Format of the data provided by EPEX.	51
2.5	Example of the partial acceptance of a bid and of unavailable bids in EPEX data	51
2.6	Example of an iceberg bid in EPEX data	51
3.1	Summary of the value function and optimal decision for step t	59
3.2	Summary of the value function and optimal decision for step 1.	61
3.3	Summary of the value function and optimal decision for step t	61
3.4	Summary of the value function and optimal decision for step $t + 1$	61
3.5	Value function for the different options in the different intervals of s_{t+1} . . .	63
3.6	Value function for the different options in the different intervals of s_{t+1} . . .	64
3.7	Summary of the value function and optimal decision for step $t + 1$	67
4.1	The balancing capacity and marginal cost of different agents for the MDP code of section 4.5.1. Units are in [MW] for P^+ and P^- , and in [€/MWh] for C	102
4.2	Results for $(D1)$, $(D3)$ and $(D4)$ in the single-agent simulation.	103
4.3	Results for $(D2)$ for different ranges of Imb_{t-1}^t in the single-agent simulation.103	
4.4	Results for $(D5)$ for different ranges of Imb_{t-1}^t in the single-agent simulation.103	
4.5	Results for $(D6)$ for different ranges of Imb_{t-1}^t in the single-agent simulation.103	
4.6	Results for different market designs using the analytical solution.	103
4.7	Results for $(D5)$ for different ranges of Imb_{t-1}^t in the analytical solution. . .	104

Chapter 1

Introduction

1.1 Motivation

In order to reduce CO₂ emissions, the European Commission launched, in 2007, the Renewable Energy and Climate Change Package which targets sourcing 20% of EU energy consumption by renewable resources by 2020 [EC08]. Thanks to this plan, the integration of renewable energy, in Europe, has increased from 13.2% in 2010 to 18% in 2019 [Eur]. This increase is foreseen to continue with the adoption of the Climate and Energy Package [EC]. This integration of renewable resources has drastically impacted electricity markets. The random availability of renewable supply creates the need for correcting system dispatch closer to real time when the forecast of renewable supply becomes more precise. An interesting option for such corrections is to trade in the Continuous Intraday Market (CIM).

Renewable energy integration may explain the recent increase of liquidity in this market. Specifically, traded volumes in the German CIM have increased from 10 TWh in 2010 to 41 TWh in 2016 [MO18]. This market is therefore becoming an interesting option for (i) fast-moving assets, such as batteries or pumped hydro storage, that can extract value from their flexibility; (ii) renewable units that can cover their forecast errors.

Another impact of the penetration of renewable resources is the development of new market designs (scarcity pricing [Hog13], flexible ramping products [CAI11]) that can better remunerate flexible assets. The need for a better remuneration stems from the following discrepancy. On the one hand, flexible assets are needed now more than ever in order to cope with the variability of renewable production. On the other hand, the profitability of these units is limited in recent years. This has been observed in [PS17], where the authors realize a simulation of the Belgian electricity market between January 2013 and September 2014. From their simulation, they observe that combined cycle gas turbine (CCGT) units, which represent the majority of flexible assets in Belgium at present, are not profitable enough to cover their investment cost. This is confirmed by Fig. 1.1 which is sourced from the yearly market monitoring report of the Belgian regulator [CREb]. This figure illustrates the evolution of the estimated (according to the regulator) profitability of a CCGT unit in the Belgian market¹. It can be observed that the profitability in recent

¹This estimation is based on an assumed trading strategy that trades 30% of the unit in the forward market (if the price is higher than the marginal cost of the unit). Then, in the short term, (i) if the price is higher than the marginal cost of the unit, the remaining energy of the unit capacity is sold, and (ii) if the price is lower than the marginal cost of the unit, the capacity offered in the forward market is bought back [CREb].

years is lower than the one that was achieved before 2012.

One of the reasons for this decreased profitability may be related to the change in the merit order curve resulting from the increased penetration of renewable units, as illustrated in Fig. 1.2. Renewable units, which have a near-zero marginal cost because they do not require burning fuels, shift the merit order curve to the right. This places flexible assets, which are often characterized by a high marginal cost, out of the market [Pap].

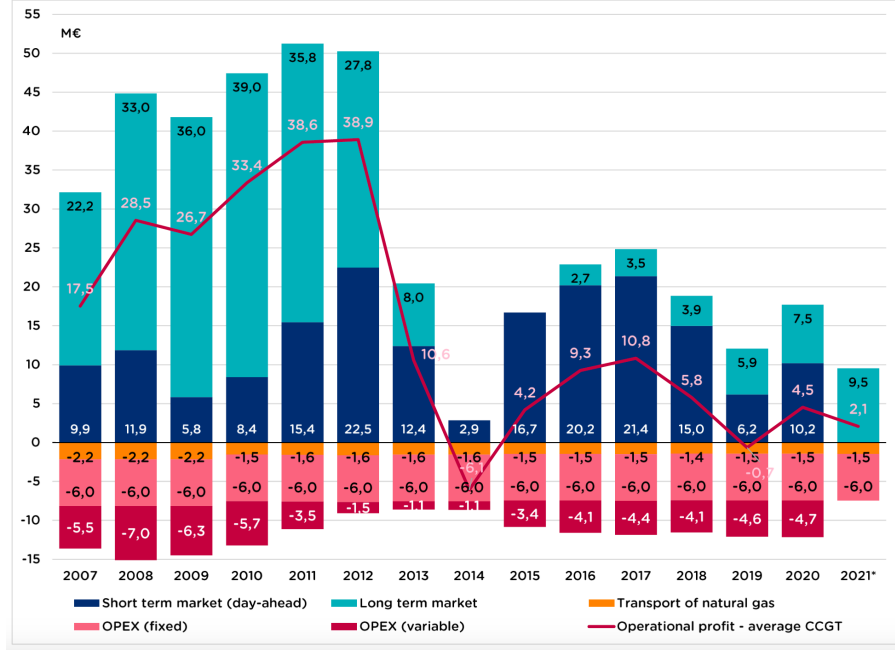


Figure 1.1: Annual profitability of a CCGT unit of 400 MW in Belgium with an efficiency of 50% and a yearly fixed cost of 7.5 million euros. The production is offered in the forward market, as well as in short-term markets [CREb].

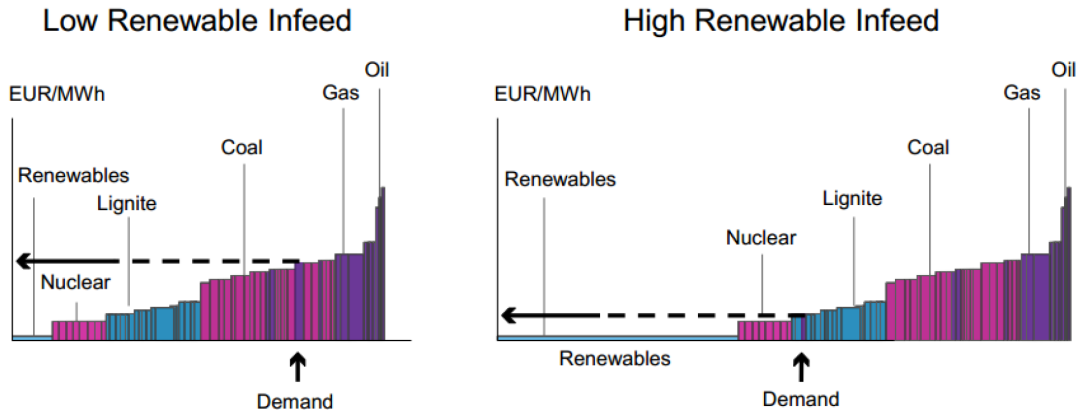


Figure 1.2: Merit order curve with low renewable infeed (left) and high renewable infeed (right) [PS].

This lack of profitability, which is observed in models, has been confirmed in practice by the fact that, in 2014, the closure of some CCGT units was announced [CREc]. As a short-term plan, in order to keep these units running, Belgium has been implementing a strategic reserve [ELIg]. The idea of this mechanism is to *keep power plants that have announced their plan to close available during winter months* [HDV16]. This mechanism will soon

be replaced by a capacity remuneration mechanism [ELIf], which aims at remunerating market participants for their installed capacity.

Another option that has been investigated by the Belgian regulator in order to improve the profitability of CCGT units is to introduce a scarcity pricing mechanism in Belgium. The idea of this mechanism is to apply an adder to real-time prices which represents the scarcity in the system (the less real-time reserve the system has, the greater the adder is). In order to study scarcity pricing, the Belgian regulator has assigned a number of studies to the Center for Operations Research and Econometrics of UCLouvain. One important conclusion of the early studies is that, if we want the mechanism to be efficient (improve significantly the profitability of CCGT units), we need the effect to not be limited only to real-time markets (which corresponds only to a small fraction of the market) but also to impact forward markets [PSB18]. This effect is referred to as **back-propagation**.

Motivated by these challenges, this dissertation focuses on intraday and balancing markets as means of enhancing the value of flexible assets. (i) **Intraday markets**. The focus in this part of the work is in developing trading strategies at high frequency for the CIM that outperform methods traditionally used in the industry. We mainly focus on three elements when developing our trading strategies: (a) We account for the impact of the trading frequency. Most methods developed in the literature consider at best hourly trading, which does not represent accurately the evolution of the CIM. The reason being that, as shown later in Table 2.1, if we trade at an hourly frequency, we only observe a fourth of the offers that are placed in the market, because the three other fourths of the offers have appeared and disappeared before we reach a decision. (b) We aim for methods that are capable of reaching decisions rapidly, since interesting offers can disappear at any moment in the CIM. (c) In order to develop an approach that can be useful for practitioners, we also aim at developing trading strategies that are based on insights that originate from analytical solutions and from detailed market analysis. (ii) **Balancing markets**. Our goal in this part of the work is to compare the performance of different market designs in order to introduce scarcity pricing in the Belgian electricity market. More specifically, we are interested in testing the ability of different options of EU balancing market designs to back-propagate the value of reserve capacity to the day-ahead reserve market. This is an important characteristic of a market design, because this back-propagation contributes towards producing a stable price signal for investment.

The remainder of this chapter is devoted to the presentation of notions that will be used through this dissertation. Section 1.2 present the concept of flexibility. Section 1.3 provides an overview of electricity market operations in Europe. Section 1.4 introduces the basis of the methodology employed in this dissertation.

1.2 Flexibility

In this section, we provide the background for understanding the concept of flexibility in electricity markets. We define flexible assets as the ones that can participate in the CIM or balancing energy market. This means that flexible resources are able to update their output at short notice. The relevant timelines that we consider in this work are the following: 30 minutes intervene between the closure of the CIM and product delivery, 3 minutes is the maximum amount of time between the activation notification and the activation start in the balancing market [ELIc].

As an example, flexible assets include storage units, combined cycle gas turbines, or

demand response units.

Another way to understand the concept of flexibility is based on option theory. There is a long literature about the use of option theory for estimating the value of power plants [DO03,DJS99]. The intuition is that owning a power plant provides the opportunity but not the obligation to burn fuel in order to produce electricity at any point in time during the lifetime of a power plant. If the owner decides to produce electricity, its payoff is equal to the spark spread (the electricity price minus the fuel cost multiplied by the heat rate). Therefore, the owner of the power plant will only operate its power plant if the spark spread is positive [DO03]. On the other side, a spark spread call option is an option the payoff of which is equal to the spark spread if it is positive or zero else [DJS99]. Therefore, it can be observed that the payoff of the power plant owner is exactly the same as the one that would be obtained from a spark spread call option. Finally, the value of the power plant can be computed as the sum of the spark spread call option payoffs over its lifetime.

The increased penetration of renewable units will influence the expected value of flexible power plants in two ways: (i) The fact that renewable units push flexible assets with high marginal costs out of the market may decrease the electricity price. This would decrease the value of the option, and therefore the expected value of owning the power plant. (ii) The increased volatility in the electricity price will increase the value of the option, because it increases the probability of the call option to be exercised. Therefore, this will increase the expected value of the power plant.

Nevertheless, this intuition based on option theory lacks two important factors: (i) Computing properly how much energy a generator can produce and how much it costs to operate it requires the modeling of important complicating features. These include ramp constraints, non-constant marginal cost, start up costs that are dependent on the time at which the power plant has been stopped, and so on. Thus, a market simulation model is a more accurate option for quantifying the value of power generating units [PS17]. (ii) It is too simplistic to treat investment decisions on the basis of expected profit alone, because investors are risk averse. Increased volatility will also discourage investments, because risk-averse investors are not willing to build power plants just based on infrequent spikes on which they might recover their investment costs. As explained in section 1.3.6, scarcity pricing is also a remedy to this risk aversion problem because, under tight conditions without load shedding, the mechanism would cause moderate but frequent price spikes of medium height. This strongly decreases the risk faced by investors, because, if the power plant is not available (e.g. forced outage) during a price spike, the investor knows that other opportunities are likely to occur for recovering the investment cost of the plant.

1.3 Overview of Electricity Markets

In this section, we provide basic background on the operation of electricity markets. We commence by presenting the positioning of each market in a timeline. We then describe each market briefly. In Fig. 1.3, we present the timing of the different short-term electricity markets in the central European power exchange. We use indicative values from the German and Belgian market. Short-term market operations commence with the balancing capacity market at 10AM the day before electricity delivery (D-1) [ELIa]. The day-ahead auction follows, with gate closure taking place at 12 noon, on D-1 [EPEf]. Subsequently, the intraday *auction* gate closure is at 3PM on D-1 [EPEf]. Following the conclusion of

the intraday auction, the continuous intraday market (which is a *separate* process from the intraday auction) commences at 3 PM for hourly products and at 4 PM for quarter-hourly products [MO18]. The CIM closes 30 minutes before delivery². The balancing energy market gate closure is 45 minutes before delivery [ELIh]. Finally, imbalances are cleared at the imbalance price [TEN].

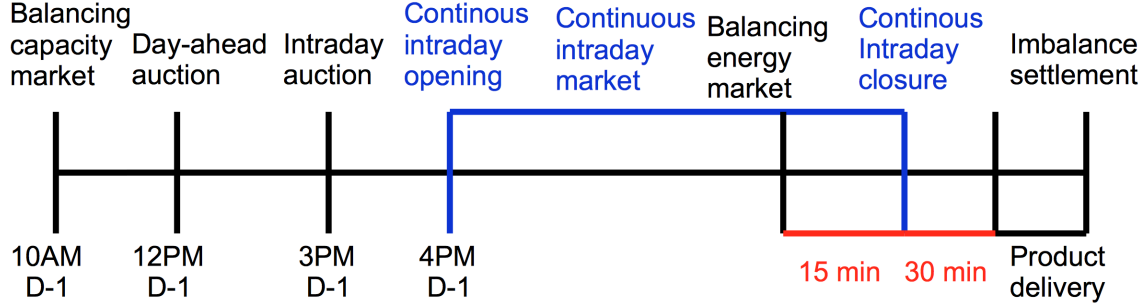


Figure 1.3: The sequence of operations in a typical central European short-term electricity market.

1.3.1 Day-ahead Auction

The day-ahead auction is a central operation of European electricity markets. The day-ahead auction is characterized by high liquidity. The total quantity traded in the day-ahead market was 234 TWh in Germany in 2016 compared to 36 TWh in the continuous intraday market and 5 TWh in the intraday auction [EPED]. Many different products can be traded in the day-ahead auction (block orders, linked blocks, exclusive blocks) [EPEg]. This allows traders to represent with a fair degree of accuracy the operational constraints of their power plants. This market is an auction, with traders submitting bids to the market. The market operator collects these bids, clears the market and announces the cleared bids and the uniform price for each delivery hour.

1.3.2 Intraday Auction

As in the case of the day-ahead auction, the intraday auction allows for a large variety of products to be traded. The granularity of the products in the intraday auction is 15 minutes, compared to the 1-hour granularity of the day-ahead market. This is interesting for traders, because it allows them to trade at the granularity that corresponds to the settlement of electricity in real time [EPEa] and [EPEe]. This is especially useful for solar units, the production of which can vary substantially within the time span of a single hour in the morning or evening. The bidding and clearing process of the intraday auction is similar to that of the day-ahead auction.

1.3.3 Continuous Intraday Market

This market is very important for renewable units because they face considerable supply uncertainty, and therefore stand to gain by adjusting their position dynamically in the CIM, as more accurate forecast information arrives for their real-time supply. This market is therefore becoming an interesting option for fast-moving assets that can valorize their flexibility by covering the uncertainty that stems from renewable production. The

²This is an indicative value for our dataset, which covers 2015 and 2016. It has been changed to 5 minutes in 2017 [KKP20]

operation of this market is drastically different compared to the one of the day-ahead and intraday auction because it is not cleared at one moment in time. Instead, it is a continuous process. At any moment of operation of the CIM, there is an order book which collects all the available bids. These bids have 4 characteristics: (i) a delivery time, which is the moment at which the power should be injected to or withdrawn from the grid; (ii) a type (sell/buy): a sell (resp. buy) bid corresponds to an offer from a counter-party to sell (resp. buy) power; (iii) a price (in €/MWh); and (iv) a quantity (in MWh). At any moment, a trader can place a new bid that is added to the order book, or accept a bid that is already present in the order book.

1.3.4 Balancing Markets

The balancing of the system in the European market design is somewhat decentralized relative to other designs that are encountered, such as the US Standard Market Design. Concretely, the balancing of the system in the European design is, to a certain extent, “outsourced” to market participants. The entities that are responsible for keeping their portfolios balanced in real-time are the Balancing Responsible Parties (BRPs), supported by the transmission system operator (TSO) who is responsible for handling any residual imbalances. BRPs are essentially portfolio owners that are responsible for ensuring that their production or consumption follows the schedule that is derived from their trades in the day-ahead auction, intraday auction and continuous intraday market, as well as any other forward markets. Nevertheless, it is possible that BRPs fail to fully balance their portfolio. This creates an imbalance between production and consumption which can cause technical problems³. In order to mitigate this imbalance, the TSO activates balancing capacity⁴. This balancing capacity is offered by balancing service providers (BSP) as follows: the capacity is committed in the balancing capacity market, and is activated in the balancing energy market⁵. Each BSP must be attributed to at least one BRP portfolio, as foreseen in article 18(4).d of the European Balancing Guideline [Eur17].

1.3.4.1 Balancing Capacity Market

In the balancing capacity market, the BSPs can offer a certain capacity of their assets for the purpose of resolving real-time imbalances. This means that, if BSPs place successful bids in balancing capacity auctions, they are required to keep this capacity available in real time in case it is requested by the TSO. In order to keep this capacity unused, BSPs receive a payment no matter if they are activated in real time or not [ELIh].

1.3.4.2 Balancing Energy Market

The balancing energy market remunerates BSPs if they are activated in real time. There are two options for bidding in the balancing energy market. According to the first option, BSPs that are cleared in the balancing capacity market for a quantity q are required to bid

³If the imbalance is too large, the electrical frequency will move too far from the reference frequency. This can cause power plants to shut down and therefore lead to a blackout.

⁴Notice that different balancing capacity products exist. They differ by the amount of time in which they should be available after notification by the TSO. In this thesis, we consider only one generic balancing capacity product. Nevertheless, the case with several balancing capacity products has also been considered in [PSB18].

⁵Balancing capacity and balancing energy are the European equivalent of day-ahead reserve and real-time energy in the US, respectively. In this thesis, we use the European and US terms equivalently.

this quantity q in the balancing energy market, but can choose the associated price [ELIh] at which they can be activated. Otherwise, if a BSP has not been cleared in the balancing capacity market and has leftover flexibility in its portfolio, it can submit an offer to the TSO for the quantity that it wishes to make available and the associated price. These are referred to as *free bids*. Thus, what distinguishes free bids from BSPs that are cleared in the balancing capacity market is that the former are not obliged to bid any capacity in the balancing energy market, whereas the latter are required to bid at least the balancing capacity that they have successfully sold in the balancing capacity market. The TSO collects all free bids, aggregates them with the aforementioned balancing energy bids, and activates them in real time according to their merit order⁶, as shown in figure 1.4 [ELId]. In this figure, the blocks correspond to the bids submitted by all the BSPs (the block height is the bid price and the block width is the quantity of the bid). These bids are ranked from the cheapest to the most expensive one. The red line represents the system imbalance which is the quantity that the TSO needs to activate to restore the system balance. The offers in blue are the ones that would be activated by the TSO. The price of the bid that is partially accepted⁷ is the balancing price that will be paid to all the activated BSPs⁸.

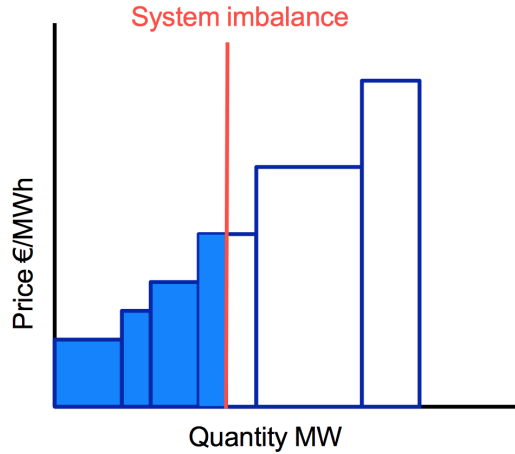


Figure 1.4: Illustration of a merit order curve.

1.3.5 Imbalance Settlement

In real time, the TSO observes the system-wide imbalance and activates balancing energy in order to address this imbalance. BRPs are settled for their real-time energy deviations at a so-called imbalance price. The settlement rules (BRP obligation and payments) faced by BRPs typically differ among different European countries, and the resulting incentives for BRPs to deviate from balancing their portfolio vary according to these different pricing rules. In Belgium, the imbalance price is computed as the price of the marginal balancing capacity that is activated, to which a term α is added in case of large system imbalance.

⁶The bids are activated from the cheapest to the most expensive one.

⁷In case the volume activated by the TSO happens to activate partially an indivisible bid, the TSO will select the next acceptable bid [ELIc].

⁸Although certain European national balancing markets currently rely on pay-as-bid settlements, the new European balancing platforms (“MARI” initiative and “PICASSO” initiative) that are being put in place will be trading at a uniform balancing price.

The rationale of the alpha parameter [ELI19] is the following: "The alpha parameter offers an additional incentive that is applicable in the event of a structural imbalance in the Belgian control area."

1.3.6 Scarcity Pricing based on an Operating Reserve Demand Curve

The aim of scarcity pricing is to align the remuneration of flexible assets that can offer reserve services to the system with the actual value of the services that they provide. The lack of profitability that flexible assets face is related to the missing money problem, that has been exacerbated by the introduction of renewable units in electricity markets [CREa].

The missing money problem arises in electricity markets in the presence of price caps. Price caps have been put in place in electricity markets in order to cope with market power issues. However, due to price caps, electricity prices may not rise sufficiently high for certain power plants to recover their investment costs. Although lifting price caps would, under ideal conditions, largely address the missing money problem, this option faces practical obstacles. Under tight system conditions, it is difficult for regulators to differentiate between legitimate price increases that are due to exhausted capacity and the exercise of market power [CREa]. Moreover, it is not desirable for producers to rely on very few price spikes that are difficult to predict, in order to recover their investment cost.

The idea of scarcity pricing based on Operating Reserve Demand Curves (ORDC) is to produce equilibrium prices for real-time balancing capacity that correspond to the level of stress in the system. These prices also affect the equilibrium price of real-time energy. The evolution of these ORDC prices as a function of the available reserve in real time is illustrated in Fig. 1.5. It can be observed that ORDC prices are low if the amount of reserve available in the system is high, and become higher when the system is tighter. The ORDC proposed by Hogan connects the equilibrium price to the loss of load probability in the system [Hog05,Hog13], and can be expressed mathematically as follows:

$$\lambda^R(R) = \max((VOLL - MC) \cdot LOLP(R), 0) \quad (1.1)$$

where:

- *VOLL* is the value of lost load. For the implementation of scarcity pricing in Belgium, it is foreseen that the current bidding limit for the balancing energy market will be used (13500 €/MWh). This choice is based on a simulation of the Belgium market [CP21]. In this simulation, two different values have been employed (8300 €/MWh and 13500 €/MWh). In order to decide between the two options, the authors propose to compare the total system cost for the 2 different options and choose the one with the smallest total cost. This total system cost includes: (i) fuel cost; (ii) fixed cost; (iii) activation cost; and (iv) shortage cost.
- *MC* is a proxy of the marginal cost of the marginal unit in the system. This can be estimated as the price of the real-time energy market. In the current design proposed in Belgium, it would be the marginal incremental price MIP (highest price for upward activated reserve)⁹.

⁹The situation will become more complex with the introduction of the European platforms MARI and PICASSO.

- R is the amount of upward reserve available in the given time period when reserve is being activated. This can either be based on data that is metered in real time, or on the offers that are placed by market participants in the balancing energy market.
- $LOLP(R)$ is the loss of load probability as a function of reserve capacity R in the system. This loss of load probability represents the probability of load shedding given the amount of reserve that is available in real time. The loss of load probability is estimated using a Gaussian distribution. Different Gaussian distributions are used depending on the season and on the moment of the day. The mean and standard deviation of $LOLP(R)$ can be derived from historical system imbalance data.

An appealing aspect of scarcity prices based on ORDCs is that these scarcity prices become non-zero when the system is tight but there is no load shedding. This is exactly their intended function. When the loss of load probability becomes non-zero frequently (the system is tight and needs investment in extra capacity), the adder will cause moderate but frequent spikes in the electricity price which will provide a signal for investing in new capacity. This is perceived as a less risky investment environment than one which relies on rare occurrences of load shedding where the price of energy reaches $VOLL$ ¹⁰.

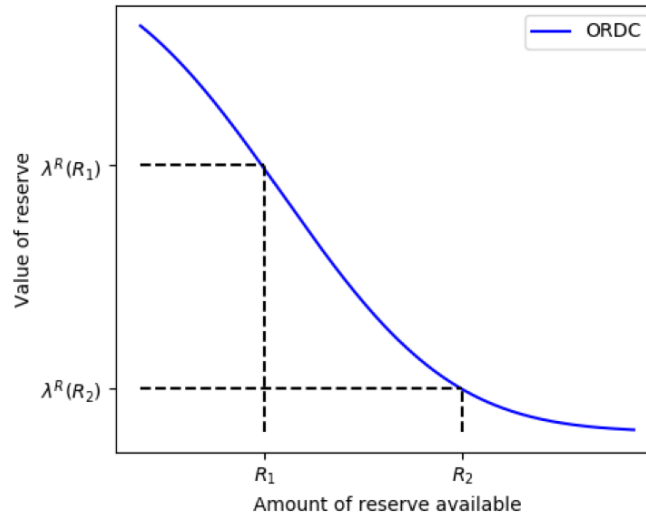


Figure 1.5: Illustration of scarcity pricing [CP21].

1.4 Reinforcement Learning

As we discuss in previous sections, both the continuous intraday market and the balancing market are useful for coping with the uncertainty of renewable supply. Therefore, the problem faced by owners of flexible assets in these markets falls under the scope of multistage optimization under uncertainty, because asset owners need to arrive to decisions while accounting for the fact that recourse actions can be adopted in an uncertain future. There are two main ways to approach this type of problem that are considered in this thesis: (i) Stochastic Programming; and (ii) Reinforcement Learning. For trading in the CIM, our method of choice is Reinforcement Learning because we are interested in

¹⁰A scarcity pricing mechanism helps in sending a price signal which is less risky for investors. Nevertheless, our work does not tackle the risk faced by investors due to regulatory uncertainty.

developing trading strategies at a high frequency. Trading at 1-second frequency would be impossible for a stochastic programming approach, because it would require solving a multistage program with several million stages. In order to compare different market designs, there are also two main approaches to the problem that can be considered (i) stochastic equilibrium methods, and (ii) multi-agent Reinforcement Learning. We opt for multi-agent Reinforcement Learning in our analysis. The reason is that, as explained in section 4.1.2, stochastic equilibrium methods are not able to model some of the features that are under consideration in the balancing market design debate.

The basic idea of Reinforcement Learning (RL) is that an agent can learn how to interact with an environment in order to achieve a goal. An agent in the context of RL is able to sense the state of the environment (state), take decisions that influence the environment (action), and quantify how well it is performing with respect to its goal (reward).

1.4.1 Markov Decision Processes

RL problems can be formalized mathematically using the framework of Markov Decision Processes (MDPs). We commence this section by defining MDPs. Subsequently, we present important notions related to the solution of MDPs. The detailed MDP formulations for the problems that are the focus of this work are presented in sections 2.3, 3.2.2 and 4.2.2.

In the MDP framework, an agent interacts with the environment for a certain number¹¹ of discrete time steps: $0, 1, \dots, T-1, T$. These interactions at each time step are illustrated in Fig. 1.6. The agent takes an action A_t . Based on this action and the previous state S_t , the environment returns a scalar reward R_{t+1} and a new state S_{t+1} . Therefore, in order to model a problem in the MDP framework, we define the following for every time step t :

- a state space \mathcal{S} which is the collection of all the information that is required in order for the agent to reach a decision.
- an action space \mathcal{A} which contains all the feasible actions of the agent.
- a reward function \mathcal{R} :

$$\mathcal{R}_s^a \doteq \mathbb{E}[R_t | S_{t-1} = s, A_{t-1} = a] \quad (1.2)$$

This represents the expected reward that is obtained if the agent chooses action a in state s .

- a transition function \mathcal{P} :

$$P_{ss'}^a \doteq \Pr[S_t = s' | S_{t-1} = s, A_{t-1} = a] \quad (1.3)$$

This provides the probability of reaching state s' if the agent chooses action a in state s . This transition function is assumed to respect the Markov property:

$$\Pr(S_t | S_{t-1}, A_{t-1}) = \Pr(S_t | S_{t-1}, A_{t-1}, \dots, S_1, A_1)$$

which means that the future evolution of the environment only depends on the current state and action, and not on past states and actions.

¹¹In this thesis, we only consider problems with a finite horizon. This implies that the agent interacts with the environment for a fixed number of time steps [SB18]. Therefore, we only describe the MDP framework and the solution methods adapted to finite horizon problems. The precise horizons for the different problems we consider are described in sections 2.3, 3.2.2 and 4.2.2.

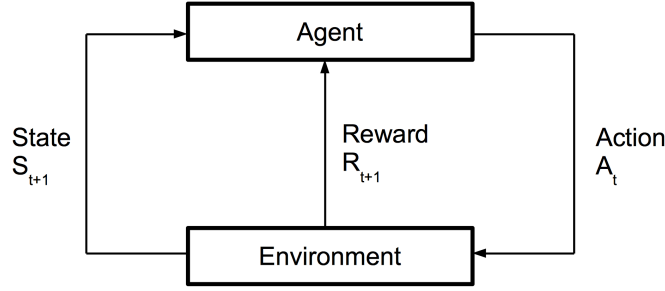


Figure 1.6: Description of the interaction between the agent and the environment in the Reinforcement Learning paradigm¹².

In order to complete the characterization of an MDP, we define the objective of the agent and what it aims at optimizing:

- The objective of the agent is to maximize the expectation of the return G_t . This return is defined as the sum of rewards until the end of an episode:

$$G_t = R_t + R_{t+1} + \dots R_T$$

- In order to maximize the expected return G_t , the agent optimizes over a set of decision rules that are followed in any state. This set of rules is referred to as a policy. A policy $\pi(a|s) = \Pr(A_t = a|S_t = s)$ is a function which is a distribution over actions for every state of the MDP¹³.

For a given initial state s_0 , the MDP problem can therefore be expressed as:

$$\max_{\pi \in \Pi} \mathbb{E}_{\pi} [G_0 | s_0], \quad (1.4)$$

Here, \mathbb{E}_{π} refers to the expectation given that an agent follows policy π .

In order to verify that a policy is optimal, we can use the Bellman optimality equation. This equation relies on the concept of a value function.

Intuition of value functions. The idea of a value function is to inform the agent about the prospect of finding itself in a given state. Concretely, the value function of a state s under policy π , denoted as $v_{\pi}(s)$, is defined as the expected return that the agent can obtain if it applies policy π starting in state s . It can be expressed as:

$$v_{\pi}(s) = \mathbb{E}_{\pi}[G_t | S_t = s] \quad (1.5)$$

Similarly, it is also possible to define the action-value function of state s and action a under policy π : $q_{\pi}(s, a)$. This is the expected return that the agent achieves if it applies action a starting in state s , and then applies policy π . It can be expressed as:

$$q_{\pi}(s, a) = \mathbb{E}_{\pi}[G_t | S_t = s, A_t = a]$$

¹²This graph is inspired from the one presented in [SB18].

¹³This is the standard definition of an MDP. In chapter 2 and 3, we employ non-deterministic policies for learning, in order to ensure sufficient exploration. Exploration refers to the notion that every potential actions should be attempted regularly, in order to check whether the decision-maker can improve on what is currently perceived as the best possible action. Nevertheless, we apply a deterministic policy when we are testing our derived policy out-of-sample.

Value function computation for a given policy. In order to compute the value function associated to a given policy, we can use the Bellman equation. The idea of the Bellman equation is to split the return between the direct reward and the return of the next state, as shown in Eq. (1.6). By applying the definition in Eq. (1.2), (1.3) and (1.5), we obtain Eq. (1.7). This shows that the Bellman equation is a set of linear equations. This means that, in order to evaluate a policy, we simply need to solve the following linear system:

$$v_\pi(s) = \mathbb{E}_\pi[R_{t+1} + G_{t+1} | S_t = s] \quad (1.6)$$

$$= \sum_{a \in A} \pi(a|s) \left(\mathcal{R}_s^a + \sum_{s' \in S} \mathcal{P}_{ss'}^a v_\pi(s') \right) \quad (1.7)$$

Similarly, the Bellman equation can also be written for the action-value function:

$$q_\pi(s, a) = \mathbb{E}_\pi[R_{t+1} + G_{t+1} | S_t = s, A_t = a] \quad (1.8)$$

$$= \mathcal{R}_s^a + \sum_{s' \in S} \mathcal{P}_{ss'}^a \sum_{a' \in A} \pi(a'|s') q_\pi(s', a') \quad (1.9)$$

Bellman optimality equation. The Bellman optimality equation is the application of the Bellman equation (1.7) for the optimal policy. It can be written as:

$$v_*(s) = \max_a \mathbb{E}[R_{t+1} + v_*(S_{t+1}) | S_t = s, A_t = a] \quad (1.10)$$

and similarly for the action-value function:

$$q_*(s, a) = \mathbb{E}[R_{t+1} + \max_{a'} q_*(S_{t+1}, a') | S_t = s, A_t = a] \quad (1.11)$$

These equations state that the value of a state under an optimal policy must be equal to the expected return assuming the best action in that state. Unfortunately, due to the max operator, Eqs. (1.10) and (1.11) are systems of non-linear equations. It is for this reason that iterative methods have been developed in order to find optimal or high-quality policies. We will present some of these methods in the following section.

1.4.2 Methods for Solving MDPs

We organize this section by first covering early methods that were developed in the literature and discussing their limitations. We then introduce more recent methods that are applicable for our problems.

Policy iteration One of the early methods that have been proposed for solving MDPs is policy iteration. This algorithm can be decomposed into two steps that are applied in alternating order:

- **Evaluate the policy:** To this aim, we need to solve the Bellman equation. This can be accomplished by solving the linear system (1.7).

- **Improve the policy:** It has been proven [SB18] that by choosing the optimal action after one step of look-ahead (1.13) according to v_π , we obtain π' which is a better policy than π .

$$\pi'(s) \doteq \operatorname{argmax}_a q_\pi(s, a) \quad (1.12)$$

$$= \operatorname{argmax}_a \mathbb{E}[R_{t+1} + v_\pi(S_{t+1} | S_t = s, A_t = a)] \quad (1.13)$$

Note that, when the policy stops changing, we are guaranteed to have converged to the optimal solution, because Eq. (1.13) becomes identical to the Bellman optimality equation (1.10).

The limitation with this method is that it relies on knowledge of the transition function as well as the reward function. For many problems, this assumption is too strong, instead we may be limited to access to samples of the transition and reward function.

Tabular methods When we can only access samples of the transition and reward function, we can resort to tabular methods. The idea of these methods is to compute a table that stores an estimation of the action-value function for all pairs of states and actions. Using this idea, the policy π can be optimized using the following algorithm [SB18]:

Initialize:

$\pi \leftarrow$ an arbitrary policy

$Q(s, a) \leftarrow$ an arbitrary action-value function

$N(s, a) \leftarrow 0$

Repeat forever:

(a) Generate an episode using $\pi : S_0, A_0, R_1, \dots, S_{T-1}, A_{T-1}, R_T$

$G \leftarrow 0$

Loop for each step of the episode, $t = T - 1, T - 2, \dots, 0$:

$G \leftarrow G + R_{t+1}$

(b) $N(S_t, A_t) \leftarrow N(S_t, A_t) + 1$

$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \frac{1}{N(S_t, A_t)}(G - Q(S_t, A_t))$

(c) $A^* \leftarrow \operatorname{argmax}_a Q(S_t, a)$ (1.14)

For all $a \in \mathcal{A}$:

$$\pi(a|S_t) \leftarrow \begin{cases} 1 - \epsilon + \frac{\epsilon}{|\mathcal{A}|} & \text{if } a = A^* \\ \frac{\epsilon}{|\mathcal{A}|} & \text{if } a \neq A^* \end{cases}$$

This algorithm can be decomposed into three steps. The first step is to run an episode following policy π . The second step is to update the value function based on the results obtained from that episode. The action-value function $Q(s, a)$ is equal to the average of all the returns that are collected after selecting action a in state s . The third step is to update the policy based on the value function. In this update, a high probability is assigned to the action that is perceived as the best and a small probability is assigned to

all other actions. This is known as an ϵ -greedy policy and is used in order to ensure that all actions are tested regularly.

The limitation of tabular methods is that it requires the state and action space to be relatively low-dimensional. Indeed, if the state/action spaces are high-dimensional or continuous, the number of states/actions can become intractable. If the number of states becomes too large, the number of times each state is visited becomes smaller, and therefore the estimates of the value function become less reliable. If the number of actions becomes too large, the algorithm becomes inefficient. This stems from (1.14), which requires finding the best action associated to an action-value function. If there are too many actions to choose from, this operation cannot be performed efficiently.

1.4.3 Approximation Methods

There are two main ideas in the literature in order to cope with high-dimensional state/action spaces. The first one is value function approximation. This methodology solves the problem of large state spaces but not the one of large action spaces. The second one is policy function approximation which can handle both large action and state spaces. After presenting these two methodologies, we describe some weaknesses that need to be considered when using these methods.

Value function approximation The idea of value function approximation is similar to that of tabular methods. The main difference resides in the estimation of the action-value function. Instead of recording the action-value function for every state-action pair $Q(s, a)$, the idea of value function approximation is to parametrize $q^*(s, a)$ as $\hat{q}(s, a; w)$, where w is a set of parameters that need to be optimized. This change does not affect the first and third step of the algorithm for tabular methods. In the second step, the same idea is followed, which is to improve the estimation of the action-value function. This can be achieved by minimizing the error between the prediction from the action-value function $\hat{q}(S_t, A_t; w_t)$ and the return from the episode G_t . This error can be expressed as:

$$L = (G_t - \hat{q}(S_t, A_t; w_t))^2.$$

We minimize this loss function using one step of a stochastic gradient algorithm. This yields the following update for w :

$$\begin{aligned} w_{t+1} &= w_t - \frac{1}{2} \alpha_t \nabla [G_t - \hat{q}(S_t, A_t; w_t)]^2 \\ &= w_t + \alpha_t [G_t - \hat{q}(S_t, A_t; w_t)] \nabla \hat{q}(S_t, A_t; w_t). \end{aligned}$$

This method can solve problems with very large state spaces. Nevertheless, the method can still be inefficient for problems with large action spaces because the algorithm requires finding the action associated with the best action-value function (line (1.14) of the algorithm).

Policy function approximation The idea of policy function approximation is to directly compute the policy without considering a value function. Specifically, policy function approximation parametrizes the policy $\pi_\theta(a|s)$ with respect to a parameter vector θ , and optimizes over this θ :

$$\pi_\theta(a|s) = \mathbb{P}[A_t = a | S_t = s; \theta].$$

An algorithm that can be employed in order to optimize the parameter θ is the REINFORCE algorithm:

- Initialize θ
- for each episode $\{S_0, A_0, R_1, \dots, S_{T-1}, A_{T-1}, R_T\} \sim \pi_\theta(a|s)$
for $t = 1 : T-1$ do

$$\theta = \theta + \gamma_t \nabla_\theta \log(\pi_\theta(a|s)) G_t \quad (1.15)$$

end
end

The REINFORCE algorithm adapts the parameter vector θ so as to maximize expected rewards from a certain policy, based on repeated episodes of the decision process. When an episode is finished, we update θ using Eq. (1.15). It has been proven in [Wil92] that the REINFORCE algorithm is effectively a stochastic gradient algorithm. It is therefore guaranteed to converge under standard stochastic approximation conditions for decreasing step-sizes γ_t . A variance-reducing adaptation to the REINFORCE algorithm is to subtract a baseline from the return G_t , as illustrated in Eq. (1.16).

$$\theta = \theta + \gamma_t \nabla_\theta \log(\pi_\theta(a|s))(G_t - b(s)) \quad (1.16)$$

In this equation, the $b(s)$ term is a baseline payoff. It can be any function of the state s , but not the action a . The presence of this baseline does not influence the expected value of the update, nevertheless it can reduce its variance [SB18]. A classical choice for this function is an estimate of the value function that can be obtained using the method presented in section 1.4.3. As we explain later, estimating a value function for our problem is a difficult task. Therefore, we use an alternative method described in section 2.6.2.

There are two potential sources of challenges related to methods that rely on function approximation:

- By using function approximation, we consider only a small subset of potential policies. This implies that we likely arrive to a policy that may be sub-optimal. There is no guarantee on how close our obtained policy is compared to the optimal one.
- There is no guarantee that the problem of optimizing the parameters of the policy is a convex problem. Therefore, we can only guarantee a local optimum and not a global one.

Deep Reinforcement Learning Deep Reinforcement Learning corresponds to the use of approximation methods for which the function approximation relies on a deep neural network. The advantage of Deep Reinforcement Learning is that it does not require any knowledge of a good parametrization. Moreover, neural networks can represent a much wider span of functions than tailor-made function approximation. The disadvantages are that (i) There are many more parameters, which implies that the optimization problem is more complicated. (ii) The results are more difficult to interpret, because the neural network essentially behaves like a black box. In this thesis, we have opted not to rely on deep Reinforcement Learning because the trading problems that we consider have never been studied using Reinforcement Learning and therefore it is meaningful to gain insights from simple parametrizations.

1.5 Contributions

The main contributions of this dissertation are organized into three chapters, which are summarized in the following.

1.5.1 Chapter 2

In chapter 2, we develop trading strategies for a storage unit in the CIM. We start by modeling the trading problem faced by the storage unit owner in the Markov Decision Process framework. After that, we present an approach based on policy function approximation for tackling the problem. This approach parametrizes the policy using 2 thresholds, one below which we buy power and one above which we sell power. We introduce and justify a collection of factors that can be used for adapting our trading thresholds to system conditions. We demonstrate the effectiveness of our approach by comparing it to the rolling intrinsic policy on real historical data. Our proposed approach outperforms the rolling intrinsic policy, which is commonly employed in practice for storage units, by increasing profitability by 17.8% on out-of-sample testing for a storage with perfect round-trip efficiency and by 13.6% for a storage unit with a round-trip efficiency of 81%. Finally, we analyse the results in order to explain the performance difference between our approach and rolling intrinsic. This chapter is based on the following publications:

- G. Bertrand and A. Papavasiliou, “An Analysis of Threshold Policies for Trading in Continuous Intraday Electricity Markets,” 2018 15th International Conference on the European Energy Market (EEM), Lodz, 2018.
- G. Bertrand and A. Papavasiliou, “Reinforcement-Learning Based Threshold Policies for Continuous Intraday Electricity Market Trading,” IEEE PES General Meeting, Atlanta, 2019.
- G. Bertrand and A. Papavasiliou, “Adaptive Trading in Continuous Intraday Electricity Markets for a Storage Unit,” IEEE Transactions on Power Systems, vol. 35, no. 3, pp. 2339 - 2350, May 2020.

1.5.2 Chapter 3

In chapter 3, we provide the first step in the direction of developing trading strategies for a renewable unit in the CIM. As a starting point, we consider selling a fixed quantity of power in an idealized version of the CIM. We start by modeling this problem in the MDP framework. Then, we derive the optimal trading strategy for this problem through backward induction. We use this analytical solution as a basis for developing value function approximation algorithm¹⁴ and an alternative Stochastic Dual Dynamic Programming that can trade under a more realistic set of assumptions. We test the performance of these two algorithms against our idealized CIM model and demonstrate that they both arrive to the optimal policy in a 10-step example. This chapter is based on the following publication:

¹⁴It would be possible to apply policy function approximation as in chapter 2 to solve the problem of chapter 3. Nevertheless, we decide to use value function approximation because we are not only interested in comparing the obtained profit with the one of a Stochastic Dual Dynamic Programming algorithm but also the value function obtained by both methods. Notice that the reverse argument is not true, the value function approximation method used in this chapter would not be suitable for the problem of chapter 2 because, as explained in section 1.4.3, it cannot handle a large action space.

- G. Bertrand and A. Papavasiliou, “Optimal Trading of a Fixed Quantity of Power in an Illiquid Continuous Intraday Market”, Powertech 2021.

1.5.3 Chapter 4

The European design is characterized by a missing market for real-time reserve capacity. This missing market undermines the valuation of reserve capacity, and the back-propagation of price signals to forward reserve markets that can support investment in reserves. The goal of chapter 4 is to develop a methodology that exposes the implications of this missing market. The methodology relies on analytical insights that can be derived under an assumption of price-taking behavior. These insights are validated by a simulation model which represents the European balancing market as a Markov Decision Process. The simulation model is used for validating the analytical insights and testing the ability of various balancing market design options to back-propagate the real-time value of reserve to forward reserve markets. This chapter is based on the following publications:

- A. Papavasiliou, G. Bertrand, A. Marien and J. Cartuyvels “Implementation of Scarcity Pricing without Co-Optimization in European Energy-Only Balancing Markets“ Utilities Policy (under review).
- A. Papavasiliou, and G. Bertrand “Market Design Options for Scarcity Pricing in European Balancing Markets,“ IEEE Transactions on Power Systems.
- A. Papavasiliou, Y. Smeers, and G. Bertrand, “An extended analysis on the remuneration of capacity under scarcity conditions,“Economics of Energy and Environmental Policy, vol. 7, no. 2, 2018.

Chapter 2

Optimizing Trading Strategies in the Continuous Intraday Market for a Storage Unit using Reinforcement Learning

2.1 Introduction

An important consequence of the integration of renewable resources in electricity markets is the need for correcting system dispatch closer to real time. An interesting option for such corrections is to trade in the CIM because it is the last market that operates before real-time market clearing, as shown in section 1.3. This explains the recent increase of liquidity in this market. Specifically, traded volumes in the German CIM have increased from 10 TWh in 2010 to 41 TWh in 2016 [MO18]. This market is therefore becoming an interesting option for fast-moving assets, such as batteries or pumped hydro storage, to extract value from their flexibility.

2.1.1 Literature Review

Several papers analyze the optimization of bidding strategies in different electricity markets. In [MS06], the authors consider trading in the day-ahead market and covering their position in imbalance for a wind power producer. This work has been extended in [CHR13] in which the authors also consider bidding in the intraday market. In [LSB18], the authors develop a trading strategy for a wind power producer who trades in the day-ahead market, followed by settlement in the real-time market. The authors account for the impact of the dependence between the wind production error and the real-time price on the trading strategy of the wind farm.

A certain body of the literature focuses specifically on storage units. The operation of storage units in the context of a US-style centralized unit commitment has been studied in the literature using unit commitment models such as in [OF15] and [KSW13]. Nevertheless, these models are out of scope in an EU context, where resource owners self-commit and self-schedule individual resources at the nomination stage which follows the clearing of the portfolio-based day-ahead market. In the EU context, the authors in [BJF14] focus on the interaction of trading strategies in the day-ahead market and the balancing

market, while the interaction between day-ahead and intraday auctions has been analysed in [Bra16].

The strategies developed for these markets cannot be applied directly to the CIM due to the continuous format of this market, which differs from the day-ahead auction or the intraday auction. Indeed, in auctions the producer has one chance to submit bids. Instead, in the CIM, the producer is afforded a certain amount of time in order to observe the offers submitted by other participants. Moreover, in the CIM, buy and sell prices for the same delivery time may evolve over the horizon of trading. Due to these particularities, the CIM has received separate treatment in the literature. The specific literature about the CIM can be classified into the three following categories.

(i) The first category of papers focuses on modeling the price evolution in the CIM. This includes literature that focuses on the explanatory variables for the evolution of the price [KP17], [Ziel17], and on the factors that influence the liquidity and the bid-ask spread [Bal18]. In [Kie17] the authors develop a Hawkes process for modeling the arrival of orders. A model for the simulation of the CIM based on data from the European Power Exchange is proposed in [MO18].

(ii) The second category of papers focuses on optimal trading strategies, and assumes that the intraday prices follow a given parametric model. Trading for a pumped hydro storage facility is presented in [BH16] and [ESFK18]. The first paper discusses the optimization problem of pumped hydro storage trading, where it is assumed that traders can access a forward curve. The second paper studies the problem of trading in the CIM and in the balancing market. Other papers consider solving for other assets. In [GM15], the authors consider trading in the CIM for balancing the forecast error of renewable energy. The authors assume that the intraday price follows a geometric Brownian motion. This is a classical assumption in finance [DO03]. Nevertheless, a geometric Brownian motion may not be appropriate to model electricity prices, since it assumes that the price is always positive, which is not guaranteed in electricity markets. A trading strategy for a thermal power plant is developed in [RAP16] for two different price models. The first price model is an additive Brownian motion which is further influenced by the most recent trades of the producers. In the second price model, the authors add the possibility of jumps in the price process. These jumps represent the situation in which renewable forecasts are inaccurate and are updated. The jump arrivals are assumed to follow a Poisson process. These jumps can either assume a fixed positive value δ^+ with probability p^+ or a fixed negative value δ^- with probability $1 - p^+$.

(iii) The third strand of literature focuses on developing trading strategies, without placing assumptions on the parametric distribution of the data. The authors in [SEM15] propose a heuristic trading method for wind power producers. In early work, the thesis author considers, in [BP18], the problem of virtual trading between the CIM and the imbalance settlement (taking a position in the CIM and closing it as an imbalance). The problem is modeled as a one-stage MDP, which is solved using policy function approximation. Related to MDP, in [BEPC18] the authors have modelled the problem of trading for a storage unit in the CIM using MDP. In order to solve this MDP, they rely on value function approximation. They prove that their strategy is efficient in-sample. The same authors extend their results out-of-sample in [BET⁺19]. Their strategy increases profitability compared to rolling intrinsic by 1.5% out-of-sample.

2.1.2 Contributions

The contributions of our work are the following: (i) We cast the intraday market trading problem for a storage unit in the MDP framework. (ii) We employ policy function approximation in order to arrive at a computationally tractable problem formulation. More precisely, we use a threshold policy according to which we seek a sell threshold above which we accept to sell power, and a buy threshold below which we accept to buy power. (iii) We propose a parametrization of the trading thresholds that accounts for several effects (e.g. the time before market closure, the delivery hour, the price in the intraday auction), in order to arrive at a policy that outperforms a benchmark policy referred to as *rolling intrinsic*. (iv) We analyse the results at higher trading frequency than the one considered in [BEPC18]: whereas in [BEPC18] the results are derived using a 5 minute frequency on a horizon of 2 hours, in the present publication we consider a horizon of 1 day with a frequency of 5 minutes for learning and 1 second for testing out of sample. Moreover, we demonstrate through experiments the important role of frequency on the training and evaluation of trading strategies.

2.1.3 Chapter Organization

Section 2.2 describes the assumptions that we place on the problem and how we simulate the continuous intraday market. Section 2.3 explains how to model the trading problem faced by a storage unit using the MDP framework. In section 2.4, we introduce the idea of a threshold policy, in order to arrive to a tractable problem for optimizing over policies. Section 2.5 presents the factors that we propose in order to adapt the threshold policy. In section 2.6, we present a test case which demonstrates the effectiveness of our approach on German market data, and we analyze how our proposed policy fares relative to rolling intrinsic. Finally, in section 2.7 we conclude the chapter and propose directions for further research.

2.2 Continuous Intraday Market Simulation

In our work, we consider trading only in the continuous intraday market. We are interested in the development of trading strategies for a storage asset owner. A storage unit is an especially interesting asset to consider in the context of intraday trading, since it offers the possibility to procure power from relatively cheap sell bids, store the power, and sell it back to subsequent buy bids that place a greater valuation on the procured power. We consider the following simplifications in our work:

1. The trading strategies that we develop are balanced. This implies that, at the closing time of the continuous intraday market, the position of the storage unit should be feasible. We thus adhere to German regulation [TEN], which requires that the producer only be in imbalance if this imbalance is caused by an unpredictable event¹ (forecast error, outages). Practically, this implies that if we do not have any energy stored in our reservoir, we cannot sell power and cover it in the imbalance market.

¹Note that US market operations differ in this respect due to central dispatch, which allows the system operator to actively manage resources in real time in order to increase trading surplus, as opposed to requiring them to remain in balance at all costs.

2. We only accept bids that are already present in the market, as opposed to also placing bids in the market. Adding the option of placing bids would complexify our Markov Decision Process in 2 ways: (i) we would have to add to our state all the bids that we have placed on the market at previous time steps. (ii) We would need to extend our action space in order to decide on suppressing the bids that we have placed at previous time steps. We have estimated an upper bound on the additional profit that could be obtained by also placing bids. The computation of this upper bound is illustrated in Fig. 2.1. The intuition is that (i) when we sell power by accepting bids, the best price that we can obtain is the highest buy bid on the market (in red), while (ii) when we sell power by placing bids, the best price that we can hope for is the lowest sell bid (in blue). Indeed, if we would place a bid at a higher price, the lowest sell bid would be accepted before our bid. Thus, the difference between the profit obtained while we accept bids and the profit that we could obtain by placing bids is the bid-ask-spread (the difference between the lowest sell bid and the highest buy bid). Therefore, we estimate the upper bound on the extra profit as the sum, over all delivery periods and time steps, of the product between the bid-ask-spread and the quantity that we trade. We obtain a potential additional profit of 27% at an hourly frequency and 69% at a 1-minute frequency. Nevertheless, we expect this upper bound on the potential gains of placing bids in the market to be loose, because there is no guarantee that another trader would accept our offer. If our offer is not matched, it may create a missed opportunity because the price might become less interesting in the future. Moreover, the looseness of the bound is expected to increase with trading frequency, because the number of trades increases with frequency.

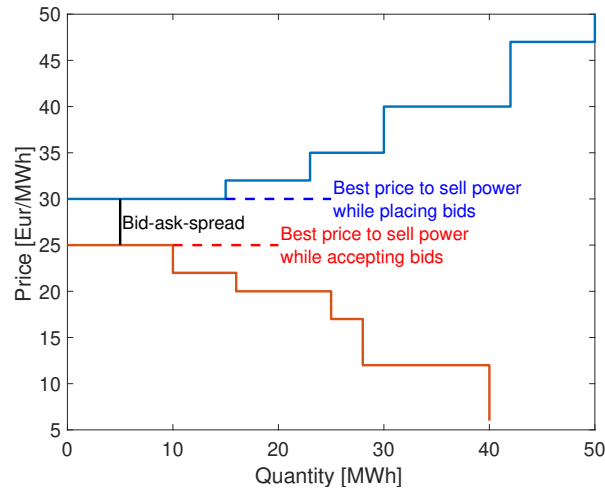


Figure 2.1: Comparison between the best price for selling power that can be achieved when accepting bids versus when placing bids.

3. In practice, CIM bids are categorized into more complex products, referred to as continuous bids, all-or-none bids (bids that cannot be partially accepted), block bids (bids that link several delivery periods), iceberg bids (bids for which the volume is split into several sub-bids that become available when the previous sub-bid has been accepted), and so on [MO18]. For our case study, we assume that all the data that we have access to corresponds to continuous bids or iceberg bids (a complete

description of how we handle the data is available in 2.8.4). This implies that we can accept fractions of bids. There are two reasons for adopting this simplification: (i) The information about the type of bids (continuous, integer, block) is not disclosed in the German market data set that we use for our case study. (ii) Practitioners have indicated to us that the impact of this restriction is minor, because most of the bids are continuous bids. To a certain extent, the more complex products have been inherited from the products that are available in the day-ahead market. A major reason for the existence of these complex products in the day-ahead market is in order to provide the option for a producer to account for complex unit commitment constraints. This interest is more limited in the CIM, because the commitment variables have to be decided several hours before delivery, through the so-called nomination procedure.

4. We only consider hourly products in our work, as opposed to also considering quarterly products that refer to delivery within a specific 15-minute interval.
5. We assume that our producer is risk-neutral. The reason for this is that the daily average profit obtained for our storage unit is around 6400 €, whereas the profit for the worst day is approximately −500 €. Typical energy companies have the financial ability to absorb this potential loss for several days without any problem. Therefore, the company can only focus on maximizing its long-term profit, which will be obtained by being risk neutral on a daily basis.
6. We assume that, no matter which bid we accept in the market, we do not influence the bids that the other actors will place later in the market. This simplification has been adopted in order to simplify the problem, and is completely in line with the state of the art on the topic of intraday trading in electricity markets [BH16, SEM15]. The only way to assess the validity of this assumption would be to run the policy in the actual market and observe the reaction of other traders to our strategy. Nevertheless, we can gain certain insights by comparing the size of our assumed asset compared to the market size. To this aim, we compare, in appendix 2.8.1, the quantity that we exchange for different delivery hours with the total exchanged quantity in the German CIM.

Using these simplifying assumptions, we describe how we simulate the evolution of the order book. To this aim, we consider 4 types of events (the complete procedure for obtaining these events based on the available market data is explained in section 2.8.4):

1. **Open**: the appearance of a trade
2. **Cancel**: the disappearance of a trade
3. **Acceptance**: the acceptance of a certain quantity of a bid
4. **Trading**: the moment when we decide which bids we accept.

The simulation of the market can now be described as follows. At the beginning of the simulation, we rank all the events, which are included in the set **Event**, chronologically. We then iterate on this set: for each new event j , we classify it in one of the 4 categories and we update the order book as described in the following procedure.

```

 $L = []$ 
for  $j \in \text{Event}$ 
  if  $j \in \mathbf{Open}$ 
    Add bid  $j$  to  $L$ 
  elseif  $j \in \mathbf{Close}$ 
    Remove bid  $j$  from  $L$ 
  elseif  $j \in \mathbf{Acceptance}$ 
    Reduce partially accepted quantity from bid  $j$ 
  elseif  $j \in \mathbf{Trading}$ 
    Launch the trading algorithm
    Remove the bids that we have accepted from  $L$ 
  end
end
end

```

2.3 Modelling the Intraday Trading Problem Using the MDP Framework

Having defined how to simulate the market, we can now analyse the trading problem. The decision problem is to decide, at different moments of the Continuous Intraday Market, which bids should be accepted in order to maximize the future expected profit of our storage unit. We consider a 1-day horizon, which is motivated by the observation that electricity prices are typically low during the night hours. Therefore, a storage unit has an interest in entering a new day with an empty reservoir and filling the reservoir up with cheap power that is available during the night hours. This makes it meaningful to decouple consecutive days, because the storage unit typically has an interest in having an empty reservoir at midnight.

In the rest of the chapter, we refer to a general storage unit. This storage unit is characterized by a certain charging and discharging efficiency. These settings create the basis for representing a battery, a simplified model of a pumped storage hydro unit, or certain types of demand response. The main trade-off for our decision problem is the following: *Do we want to trade power at the current price and lock in the profit? Or is it worth waiting for a potential future bid, the price of which would be more advantageous, despite the risk that the current favorable bids may disappear?* This problem enters the scope of Reinforcement Learning described in section 1.4, and as explained in section 1.4.1, a common way to model this class of problems is the Markov Decision Process framework.

2.3.1 State Variables

In order to reach a decision at time step t , we require 3 ingredients in our state \mathcal{S}_t : (i) The offers available in the continuous intraday market at time step t . This data is available in the market order book. (ii) A variable $v_{t-1,d}, \forall d \in D$ which indicates the capacity that would be stored in the storage unit at delivery hour d if we were only executing the trades decided at time step $t - 1$ or earlier. This value can be easily computed based on the

results of all the trades that we have realized in the past. (iii) Exogenous data that we anticipate should influence our decision. Some examples of these exogenous parameters include the remaining time before market closure, and the price of the intraday auction. The full list of these parameters, and the way in which we use them, is discussed in section 2.5.

2.3.2 Action Variables

In order to model our action space \mathcal{A}_t , we require one action variable $a_{t,d}$ for each delivery time d . This action indicates how much we wish to sell at time step t . In theory, this variable can be continuous. But, in order to reduce the size of the action space, we discretize this variable into $2n + 1$ potential actions:

$$a_{t,d} \in \{-q_n, \dots, -q_1, 0, q_1, \dots, q_n\}$$

2.3.3 Reward

The total reward obtained from the CIM at time step t is equal to the sum of the rewards obtained for every delivery hour:

$$R_t = \sum_{d \in D} \text{rev}(a_{t,d}),$$

where the reward for delivery hour d at time step t is computed as the integral of the demand curve $p_{t,d}$ from 0 to $a_{t,d}$:

$$\text{rev}(a_{t,d}) = \int_0^{a_{t,d}} p_{t,d}(z) dz.$$

2.3.4 State Transition Function

For this intraday trading problem, we do not have access to the transition function (since we do not place any assumptions on the evolution of intraday prices). As explained in section 1.4.2, this prohibits us from using methods such as policy iteration. Nevertheless, we are perfectly within the scope of Reinforcement learning, because tabular methods (see section 1.4.2) and function approximation methods (see section 1.4.3) do not require any knowledge of the transition function to be applied. Note that the round-trip efficiency of a storage unit is part of this transition function, which we do not model explicitly.

2.4 Threshold Policy

As we have to simultaneously reach decisions for the 24 delivery periods, our state and action spaces are intractable for tabular methods and value function approximation methods. Therefore, we rely on policy function approximation. More precisely, we focus on a policy which is parametrized by buy and sell price thresholds. The threshold policy that we investigate in this chapter accepts sell bids if their price is below a buy threshold, and accepts buy bids if their price is above a sell threshold. Our focus on threshold policies is justified by several factors: (i) Optimal inter-temporal arbitrage in a deterministic setting is achieved by a threshold policy if the reservoir limit constraint is not binding, as proven in appendix 2.8.2. (ii) Threshold policies have also been *proven* to be optimal in a number

of papers in the literature regarding specific instances of stochastic optimal control problems with uncertain prices [Mor59, Kin69, Gol85]. (iii) The idea of using a threshold policy in order to trade for a storage unit has already been proposed in other settings [PM16b].

We apply a stochastic² threshold policy, in order to ensure that sufficient exploration takes place during the learning stage of the algorithm. Concretely, we propose drawing the sell and buy thresholds from a Gaussian distribution. Therefore, we define our policy parameter, θ , as $\theta = (\mu_X, \sigma_X, \mu_Y, \sigma_Y)$, where the buy threshold for delivery hour d , X_d , is drawn according to a normal distribution with parameters³ $(\mu_X, \exp(\sigma_X))$, and the sell threshold for delivery hour d , Y_d , is drawn according to a normal distribution with parameters $(\mu_Y, \exp(\sigma_Y))$. We draw one threshold per delivery hour d . Therefore, the distribution of actions over all future delivery hours can be decomposed as the product of the distribution for each delivery hour. Mathematically, this can be expressed⁴ as:

$$\pi_\theta(a|s) = \prod_{d \in D} \pi_\theta^d(a_d|s) \quad (2.1)$$

In order to illustrate how the stochastic threshold is implemented, we consider the example of Fig. 2.2. at delivery hour d . (i) The green decreasing function corresponds to the buy bids that are available in the order book for delivery hour d . This data is available in the order book at the time we are deciding on whether or not to accept a bid. The demand curve is associated with the lower x-axis. (ii) The bell curve represents the probability density function of the threshold. This curve can be computed based on the current vector parameter θ . The bell curve is associated with the upper x-axis. With these two elements, we illustrate how we use the threshold policy in order to arrive at decisions. Consider, for instance, the action *Sell* 10 MWh: if the sell threshold that we draw is between the price associated to a sell quantity of 15 MWh and the price associated to a sell quantity of 5 MWh, we sell 10 MWh. The probability of this action corresponds to the red surface $\pi_\theta^d(10|s)$. This probability can also be computed mathematically, as illustrated below:

$$\begin{aligned} \pi_\theta^d(10|s) &\triangleq \Pr(a_d = 10) \\ &= \Pr(p(15) \leq Y_d \leq p(5)) \\ &= \Pr(Y_d \leq p(5)) - \Pr(Y_d \leq p(15)) \\ &= \Phi(p(5); \mu_Y, \exp(\sigma_Y)) - \Phi(p(15); \mu_Y, \exp(\sigma_Y)) \end{aligned}$$

where $\Phi(\cdot; \mu, \sigma)$ indicates the cumulative distribution function of the normal distribution with mean μ and standard deviation σ . In order to apply the REINFORCE algorithm, we also need to compute the policy derivatives for the different actions. These derivatives can be computed analytically as illustrated below for the derivative of the probability of the action *Sell* 10 MWh with respect to μ_Y :

²Notice that our threshold is only stochastic in the learning phase. When we test the performance out-of-sample, we apply deterministically the thresholds that have been learned in the training phase.

³We use $\exp(\sigma_X)$ and not σ_X directly in order to ensure that the standard deviation remains positive.

⁴In order to simplify notation, we present the situation for which the actions for the different delivery hours are independent of each other. If this is not the case, a similar formula using the conditional distributions can be used.

$$\begin{aligned}
\frac{\partial \pi_{\theta}^d(10|s)}{\partial \mu_Y} &= \frac{\partial \Phi(p(5); \mu_Y, \exp(\sigma_Y))}{\partial \mu_Y} - \frac{\partial \Phi(p(15); \mu_Y, \exp(\sigma_Y))}{\partial \mu_Y} \\
&= -\phi(p(5); \mu_Y, \exp(\sigma_Y)) + \phi(p(15); \mu_Y, \exp(\sigma_Y))
\end{aligned}$$

where $\phi(\cdot; \mu, \sigma)$ denotes the probability density function of the normal distribution with mean μ and standard deviation σ .

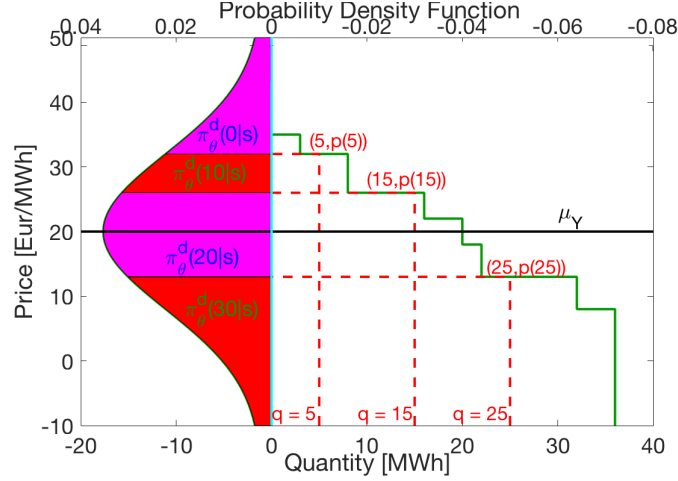


Figure 2.2: Threshold policy for the hydro problem if we consider four possible actions: sell 0, 10, 20 or 30 MWh. The bell curve indicates the probability density function of the sell threshold. The two purple segments and the two red segments of the bell curve indicate the probability of each of the four actions. The green decreasing function corresponds to the buy bids that are available in the order book.

2.5 Factors Driving the Optimal Threshold

In the previous section, we have developed a basic threshold policy for trading in the CIM. This simple threshold policy does not achieve satisfactory performance in practice, because it is not sufficient to maintain the same threshold for every time step of every day. This suggests that the threshold should be further dependent on certain factors that are pertinent towards an adaptive trading strategy. In this section, we propose a number of such factors and explain the reason for which we consider them. Then, we explain how the REINFORCE algorithm can be adapted in order to incorporate these factors.

2.5.1 Delivery Time

The need for using different thresholds depending of the delivery hour is illustrated in Fig. 2.3. This graph represents the *CIM price* (which we define as the center of the bid-ask spread) for the 24 different delivery hours. The red dot represents the price of buying energy at the 6th hour, while the green dot is the price of buying energy at the 17th hour. These two prices are equal, however the buying decision should be different. Indeed, the price corresponding to the red dot is not interesting, because the same amount of power could have been procured and stored at the reservoir at a lower price at hour 4. On the contrary, the price corresponding to the green dot is interesting, because it corresponds to a local minimum price. Thus, in hour 17 we can buy power, in order to sell that power back at a later delivery time.

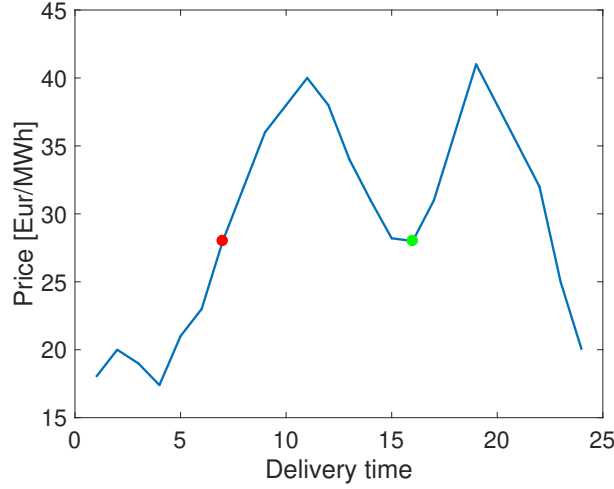


Figure 2.3: The delivery time of an order impacts its threshold: buying power at 30 €/MWh is not worthwhile in hour 6, but it is worthwhile in hour 17.

Having argued that it is necessary to employ different thresholds for different delivery times, our idea is to define regimes for which the threshold mean should be the same. We will define these regimes based on the intraday auction price curve, which conveys a significant amount of information about the CIM price.

We present an example of these regimes based on the intraday auction price curve, for one precise day of our dataset, in Fig. 2.4. The justification of why we use the intraday auction in order to compute the regimes is provided in appendix 2.8.5.1. These graphs illustrate that the buy threshold switches at the maximum of the price curve, since any power that we buy between two maxima can be sold at the second maximum. Similarly, the sell threshold switches at the minimum of the price curve, because any power that we sell between two minima can be bought at the first minimum⁵. On average, there are 1.53 regimes per day.

The introduction of regimes impacts the parameter vector θ . Since we introduce different thresholds for the different regimes, μ_X and μ_Y are now indexed by the regime k , and are thus denoted as μ_X^k and μ_Y^k . In the remainder of this section, we will express these threshold means⁶ μ_X^k and μ_Y^k as a function of 10 parameters⁷, which we denote as $(\alpha_1^s, \alpha_1^b, \alpha_2^s, \alpha_2^b, \alpha_3^s, \alpha_3^b, \alpha_4^s, \alpha_4^b, \alpha_5^s, \alpha_5^b)$. We will then show how the REINFORCE algorithm can be used in order to learn the values of the parameter vector α .

2.5.2 Intraday Auction Curve

Our motivation for using the intraday auction curve as a feature for determining thresholds is illustrated in Fig. 2.5, where we present the *CIM price* for two different trading days. From this graph it is clear that it is not possible to set a single threshold which would

⁵This reasoning is slightly simplified. The full explanation on how the regimes are computed can be found in appendix 2.8.5.2.

⁶In contrast to the mean, we do not make the standard deviation dependent on exogenous parameters. This is due to the fact that the standard deviation is only used in order to ensure sufficient exploration in the learning phase.

⁷The final thresholds μ_X^k and μ_Y^k are dependent on the regime, however the α parameters are not. Therefore, the number of regimes does not affect the number of parameters that need to be learned.

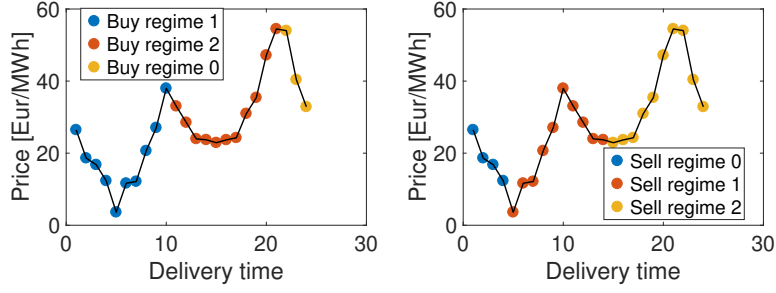


Figure 2.4: Buy regime (left), and sell regime (right) based on the intraday auction price.

perform well for both days, because the average level of the curves is different. In order to set an appropriate base level for the thresholds, we use the intraday auction price. The idea is that the price of previous markets can provide an indication about the state of the market, and thus support the forecast of the price for subsequent market-clearing stages. This observation has been inspired by (i) reference [SZ19], in which the authors use future prices in order to forecast the day-ahead market prices; (ii) reference [MRRFJC16], where the authors use the day-ahead market prices and past intraday prices in order to forecast the next intraday prices in the Spanish market⁸; (iii) reference [KEF15], in which the authors use the last balancing price in order to forecast the next balancing price; and (iv) reference [BJF14], where the authors observe a strong correlation between the day-ahead market and the balancing market.

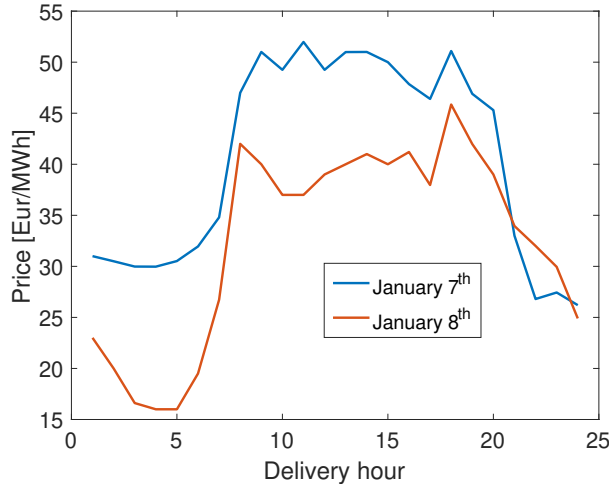


Figure 2.5: Continuous intraday market price for two different days. The curves correspond to different average values, therefore different thresholds need to be applied for an effective threshold strategy.

Motivated by this observation, we propose an adaptation of the thresholds as follows:

$$\begin{aligned}\mu_X^k &\leftarrow p_{\min,k} + \alpha_1^s(p_{\max,k} - p_{\min,k}) \\ \mu_Y^k &\leftarrow p_{\max,k} - \alpha_1^b(p_{\max,k} - p_{\min,k})\end{aligned}$$

where $p_{\min,k}$ is the minimum of the k^{th} buy regime of the intraday auction curve, $p_{\max,k}$ is the maximum of the k^{th} sell regime of the intraday auction curve, and $\alpha_1^{s/b}$ are the weights

⁸Note that, in Spain, there is no continuous intraday market but rather 6 sequential auctions.

that will be optimized using the REINFORCE algorithm.

The idea behind this parametrization is that $p_{\min,k}$ (resp. $p_{\max,k}$) is a reasonable starting point for the buy (resp. sell) threshold because it is the best price that could have been obtained in the intraday auction for regime k . Then, with the parameter α_1^s (resp. α_1^b) we allow the REINFORCE algorithm to determine to what extent the threshold should move from $p_{\min,k}$ (resp. $p_{\max,k}$) to $p_{\max,k}$ (resp. $p_{\min,k}$), based on learning from repeated episodes. In order to apply the REINFORCE algorithm for learning the parameter vector α , we need to compute the derivative of our policy with respect to α . The derivative can be computed using the chain rule, as we show in Eq. (2.2) for the derivative of α_1^b , for delivery hour d in regime k .

$$\begin{aligned} \frac{\partial \pi_\theta^d(a_d|s)}{\partial \alpha_1^b} &= \frac{\partial \pi_\theta^d(a_d|s)}{\partial \theta}^T \frac{\partial \theta}{\partial \alpha_1^b} \\ &= \frac{\partial \pi_\theta^d(a_d|s)}{\partial \mu_Y^k} (p_{\min,k} - p_{\max,k}) \end{aligned} \quad (2.2)$$

2.5.3 Quantity Already Traded

The intuition for this adaptation is that, at any stage of the trading process, if we have already bought a large quantity of power and have not sold it yet, we wish to avoid the risk of ending up with unsold power. Note that we assume that there is no residual value for leftover water in the reservoir at the end of the horizon, which is consistent with the fact that we have an interest in entering a new day with an empty reservoir and filling the reservoir up with cheap power that is available during the night hours.

In order to capture this effect, we add a penalty in order to accept buying at a lower price and to accept selling at a lower price:

$$\begin{aligned} \mu_X^k &\leftarrow \mu_X^k - \alpha_2^s \cdot v_{\text{end}} \\ \mu_Y^k &\leftarrow \mu_Y^k - \alpha_2^b \cdot v_{\text{end}} \end{aligned}$$

where v_{end} is the volume that we would obtain at the last delivery period with the trades that we have already engaged in. This adjustment of the thresholds implies that, moving forward, we become less selective about selling power and more selective about buying power, until the reservoir eventually becomes empty.

2.5.4 Remaining Time before Market Closure

Whenever the producer has not sold all the energy stored in its reservoir close to the maximum of a regime, the producer should become less selective in the price it asks. This is due to the fact that there are few subsequent opportunities to trade, and the currently observed price is possibly the best price that the producer can secure for the trade. Similarly, whenever the producer has not bought up to the capacity of its reservoir as it is approaching the minimum of a regime, it should become less selective with the price that it asks for buying power. This approach is inspired by the theory of the optimal stopping problem [Lin61], [TR99].

We capture this effect by varying the threshold means as follows:

$$\mu_X^k \leftarrow \mu_X^k + \alpha_3^s \frac{p_{\max,k} - p_{\min,k}}{2} \exp(\alpha_4^s (t - T_k^s)) \quad (2.3)$$

$$\mu_Y^k \leftarrow \mu_Y^k - \alpha_3^b \frac{p_{\max,k} - p_{\min,k}}{2} \exp(\alpha_4^b (t - T_k^b)) \quad (2.4)$$

where t is the current time step, T_k^b is the delivery time of the maximum of the k^{th} sell regime, and T_k^s is the delivery time of the minimum of the k^{th} buy regime.

We employ 4 coefficients in Eqs. (2.3) and (2.4): (i) $\alpha_3^{s/b}$ determines the strength of this effect; (ii) $\alpha_4^{s/b}$ determines how smoothly the threshold adapts with respect to the gate closure time. A large value for $\alpha_4^{s/b}$ would decrease the selectivity very close to the delivery time. On the contrary, a small value for $\alpha_4^{s/b}$ would decrease the selectivity more smoothly with respect to the remaining time.

2.5.5 Relative Value of Observable Bids

The motivation for this factor is to account for the coupling among the bids of different delivery hours, due to the fact that the battery can only store a finite amount of energy. Concretely, we wish to avoid accepting a bid even though the order book includes a bid at an adjacent delivery period that can be traded for a better price. In order to account for this inter-dependency, we penalize the bids that would not be accepted by the rolling intrinsic method. The rolling intrinsic method is a myopic policy which means that it will select the subset of trades which can be absorbed by the reservoir without exposing the unit to imbalances, and will do so by maximizing the profit of the current time step (the rolling intrinsic method is presented in more details in section 2.6.1).

Concretely, the adjustment to our algorithm is illustrated in Fig. 2.6. The figure corresponds to the case in which rolling intrinsic sells 20 MWh for delivery period d . When this occurs, we wish to decrease the probability of selecting the action of selling 30 MWh, and reallocate it to the probability of selling 20 MWh. To this end, we introduce an auxiliary Gaussian distribution with a mean of $\mu_Y^k + \alpha_5^b$ and with a standard deviation of $\exp(\sigma_Y)$. We compute the probability of the action *Sell* 30 MWh by using a threshold drawn from the auxiliary Gaussian distribution, which is indicated with the green bell curve in the figure. Mathematically, this is illustrated in Eq. (2.8). This change decreases the probability of the action *Sell* 30 MWh, relative to the probability that would have been obtained from the original bell curve of Fig. 2.6. The difference in probability mass is transferred to the last action that is accepted by rolling intrinsic (*Sell* 20 MWh), as illustrated in Fig. 2.6 and in Eq (2.7). As we can see in the figure, the higher the value of α_5^b , the less likely we are to choose the action that is not selected by rolling intrinsic. Finally, the probability for (*Sell* 0 MWh) and (*Sell* 10 MWh) is exactly the same as in the initial situation, as shown in Fig. 2.6 and in Eqs. (2.5) and (2.6). The reason is that these actions are feasible, and do not correspond to the feasible action with the greatest volume of trading (*Sell* 20 MWh), therefore we wish to keep their probability of being selected intact.

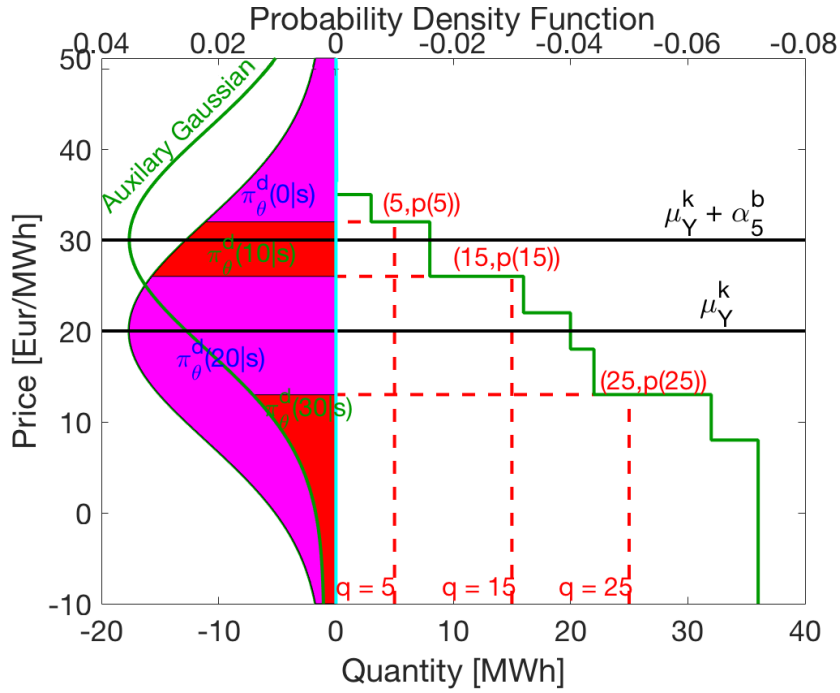


Figure 2.6: Illustration of the probability reallocation that relies on the auxiliary Gaussian distribution, as described in section 2.5.5.

- **Sell 0 MWh:**

$$\begin{aligned}\pi_\theta^d(0|s) &= \Pr(a_d = 0) \\ &= \Pr(Y_d \geq p(5)) \\ &= 1 - \Pr(Y_d \leq p(5)) \\ &= 1 - \Phi(p(5); \mu_Y^k, \exp(\sigma_Y))\end{aligned}\tag{2.5}$$

- **Sell 10 MWh:**

$$\begin{aligned}\pi_{\theta}^d(10|s) &= \Pr(a_d = 10) \\ &= \Pr(p(15) \leq Y_d \leq p(5)) \\ &= \Pr(Y_d \leq p(5)) - \Pr(Y_d \leq p(15)) \\ &= \Phi(p(5); \mu_Y^k, \exp(\sigma_Y)) - \Phi(p(15); \mu_Y^k, \exp(\sigma_Y))\end{aligned}\quad (2.6)$$

- **Sell 20 MWh:**

$$\begin{aligned}\pi_{\theta}^d(20|s) &= \Pr(a_d = 20) \\ &= \Pr(Y_d \leq p(15)) - \Pr(Y_d + \alpha_5^b \leq p(25)) \\ &= \Phi(p(15); \mu_Y^k, \exp(\sigma_Y)) - \Phi(p(25); \mu_Y^k + \alpha_5^b, \exp(\sigma_Y))\end{aligned}\quad (2.7)$$

- **Sell 30 MWh:**

$$\begin{aligned}\pi_{\theta}^d(30|s) &= \Pr(a_d = 30) \\ &= \Pr(Y_d + \alpha_5^b \leq p(25)) \\ &= \Phi(p(25); \mu_Y^k + \alpha_5^b, \exp(\sigma_Y))\end{aligned}\quad (2.8)$$

In order to implement the REINFORCE algorithm, we also need the policy derivative for the different actions. The development in order to compute these is made in Eqs. (2.9), (2.10), (2.11) and (2.12). The derivatives for 0 and 10 MWh are equal to 0 because α_5^b does not appear in the probability of these actions. The derivative for the action *Sell 30 MWh* is negative. This is coherent with the intuition that the higher the value of α_5^b , the less likely we are to choose the action that is not selected by rolling intrinsic. The derivative for the action *Sell 20 MWh* is the opposite of the one of the action *Sell 30 MWh*. This is due to the fact that the probability lost by the action *Sell 30 MWh* is exactly transferred to the action *Sell 20 MWh*.

$$\frac{\partial \pi_\theta^d(0|s)}{\partial \alpha_5^b} = 0 \quad (2.9)$$

$$\frac{\partial \pi_\theta^d(10|s)}{\partial \alpha_5^b} = 0 \quad (2.10)$$

$$\frac{\partial \pi_\theta^d(20|s)}{\partial \alpha_5^b} = \phi(p(25), \mu_Y^k + \alpha_5^b, \exp(\sigma_Y)) \quad (2.11)$$

$$\frac{\partial \pi_\theta^d(30|s)}{\partial \alpha_5^b} = -\phi(p(25), \mu_Y^k + \alpha_5^b, \exp(\sigma_Y)) \quad (2.12)$$

2.5.6 Preventing Imbalances

As we explain in the assumptions in section 2.2, we are only interested in developing trading strategies that do not result in imbalance. Therefore, in order to suppress actions that result in imbalances, we re-assign their probability to the closest action which does not result in an imbalance, using the same idea as in section 2.5.5. In this case, the parameter α_5^b is replaced by a constant M which is sufficiently large in order to ensure that an action which would result in an imbalance is never selected.

2.5.7 Adapting with Respect to Round-Trip Efficiency

In order to account for round-trip efficiency, we present an example that illustrates the concept of *perceived* efficiency, which distinguishes whether we are planning to cover a bid *financially* or *physically*. Suppose that we have two delivery hours, a charging efficiency η_{in} of 0.9 and a discharging efficiency η_{out} of 0.9. Suppose that we have already bought 20 MWh for the first delivery hour at the previous time step. Therefore, the quantity that would be stored is 18 MWh for both delivery times. If we want to sell power at the second delivery time, we can only sell 16.2 MWh, because we have to apply the discharge efficiency. We define the *perceived* efficiency for this order as $\frac{16.2}{18} = 0.9$. On the contrary, if we want to sell at the first delivery time, we can sell 20 MWh, because this operation will simply cancel the previous purchase of 20 MWh. This is a purely financial operation. We thus define the *perceived* efficiency for this order as $\frac{20}{18} = 1.11$.

In order to account for this effect in the threshold parametrization, we use the same idea as in section 2.5.5.

- We determine a certain *baseline* for the mean of the Gaussian distribution of our buy and sell threshold, which corresponds to the case in which we are accepting a certain quantity that serves as a purely *financial* transaction. We then adapt the

thresholds as follows:

$$\begin{aligned}\mu_X^k &\leftarrow \frac{1}{\eta_{\text{out}}} \mu_X^k \\ \mu_Y^k &\leftarrow \eta_{\text{in}} \mu_Y^k\end{aligned}$$

This adaptation is coherent with the intuition presented in the example. Indeed, if we are canceling a position that we have previously taken in the market, we can accept a less interesting price (i.e. accepting a lower sell / higher buy threshold), because the perceived efficiency is higher than 1.

- The mean of the auxiliary Gaussian represents the case in which we are opening a new position. Therefore, the auxiliary Gaussian distribution mean will be equal to (i) $\eta_{\text{in}} \cdot \eta_{\text{out}} \cdot \mu_X^k$ for the buy threshold, and (ii) $\frac{\mu_Y^k}{\eta_{\text{in}} \cdot \eta_{\text{out}}}$ for the sell threshold. This is also coherent with the example, because we are requesting a more selective price if we are opening a new position than if we are engaging in a purely financial transaction, since the *perceived efficiency* is less good. Note that this adaptation does not add any new parameters in the learning algorithm.

2.6 Case Study

In this section we present results from the implementation of the proposed policy on the German continuous intraday market. The data for the German CIM has been procured from the European Power Exchange (EPEX), and spans the years 2015 and 2016⁹. For the purpose of the case study, we place ourselves in the position of a storage asset owner who manages a unit with a maximum storage capacity of 200 MWh and that has not made any trades in earlier markets (day-ahead, intraday auction). We assume that, on July 19, the owner adopts our strategy and has at its disposal market data since the beginning of the year¹⁰. Therefore, we use the 200 first days of 2015 as training set, and the last 165 days of 2015 and the 366 days of 2016 as a test set.

2.6.1 Rolling Intrinsic Method

We compare our approach to the rolling intrinsic method. This method has already been used as a benchmark in the literature [BEPC18]. This method is also popular in the gas trading context [LW21].

The rolling intrinsic policy is a myopic method for trading in continuous markets. The idea of the method is to trade so as to maximize the instantaneous reward at each time step [PM16a]. In the context of our problem, the rolling intrinsic method will select

⁹There have been a number of changes in the German CIM since 2016. Firstly, the time between the closure of the CIM and the product delivery was 30 minutes in 2016, and has been changed to 5 minutes in 2017 [KKP20]. Secondly, liquidity has kept increasing in Germany. The exchanged quantity in 2016 was 41TWh, and reached 56TWh in 2020 [EPEb]. Finally, cross-border bids have been introduced in the CIM in June 2018 [EPEc]. These bids allow for trading between different countries if there is sufficient interconnection capacity available. Nevertheless, this change is less drastic for the German market, because it still represents 56% of the total exchanged volume in European CIMs in 2020 [EPEb]. These changes do not affect the possibility of applying our method. Moreover, there is no reason to believe that the results would be significantly different if the method would be applied on more recent data featuring these changes.

¹⁰Note that data which extends too far back in time may not be as useful, due to the rapid structural evolution of the market (increase in renewable energy integration, changes in market design, etc.).

the subset of trades which can be absorbed by the reservoir without exposing the unit to imbalances, and will do so by maximizing the profit of the current time step. This myopic policy can be written as an optimization problem at every time step of trade. The optimization model is developed in appendix 2.8.3.

2.6.2 Learning Process

In order to optimize our policy parameters, we rely on the REINFORCE algorithm as described in section 1.4.3. In order to reduce the update variance, we use a baseline function. It would be complicated to find a good parametrization in order to compute an estimation of the value function, as presented in section 1.4.3. Indeed, the expected profit (value function) of a storage asset is highly non-linear with respect to the offers available in the market, as it is not an offer in itself that creates a potential profit but rather a combination of offers. In order to avoid this complexity, we use the instantaneous profit that rolling intrinsic would obtain on the available offers as a baseline function.

We aim at learning the optimal threshold, so as to apply our threshold policy with a frequency of 1 second. We consider 1 second as a sufficiently high frequency for testing the algorithm in the continuous intraday market because, as observed in Table 2.1, if we trade every second, we will observe 98.3% of the offers. This means that almost all of the offers remain in the market for at least one second¹¹, before being matched with competing offers on the platform. Notice that our method can be used for trading at any frequency, because the computational effort for optimizing the policy parameters is performed off-line. Therefore, applying our threshold policy in real time is instantaneous.

In order for the learning stage of the algorithm to be computationally tractable, we gradually refine the learning frequency from hourly steps to 15-minute steps and ultimately to 5-minute steps¹², as indicated in Fig. 2.7. In this figure, 1 iteration corresponds to 4 repetitions of the 200 days of learning, which amounts to 800 episodes. These episodes are executed in parallel on an HPC cluster using 8 CPUs for 40 hours.

Length of time step	Percentage of offers observed
1 hour	25.7
15 minutes	41.5
5 minutes	56
1 minute	74.8
15 seconds	86.7
5 seconds	92.2
1 second	98.3

Table 2.1: Percentage of offers that are observed as a function of frequency of accessing the market data.

A potential issue for our learning phase is that the REINFORCE algorithm is a stochastic algorithm. Therefore, different runs can produce different results. In order to test the sensitivity of our results, we have conducted an experiment in which we run 6 different

¹¹The average time between two offer arrivals is 29.9 seconds. Nevertheless, the offer arrivals are not distributed uniformly over time. Indeed, during the last hour, the average time between two offer arrivals is 4.3 seconds.

¹²We decide to switch to a higher learning frequency when the profit appears to stabilize. This is due to the fact that there is no reason to run the algorithm until full convergence for the hourly frequency, because it is not the problem we are interested in (the order data arrives at much higher frequency than hourly).

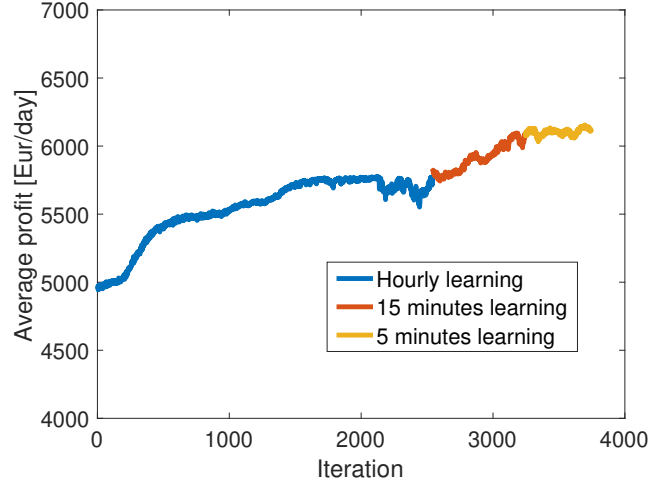


Figure 2.7: Evolution of profit as a function of iterations of the REINFORCE algorithm.

realizations of the REINFORCE algorithm at hourly frequency and compare the evolution of the alpha parameters and the profit. We illustrate the results for parameter α_4^s (left panel of Fig. 2.8) and the average profit (right panel of Fig. 2.8). We observe that the different runs exhibit very similar average profit and parameter evolution.

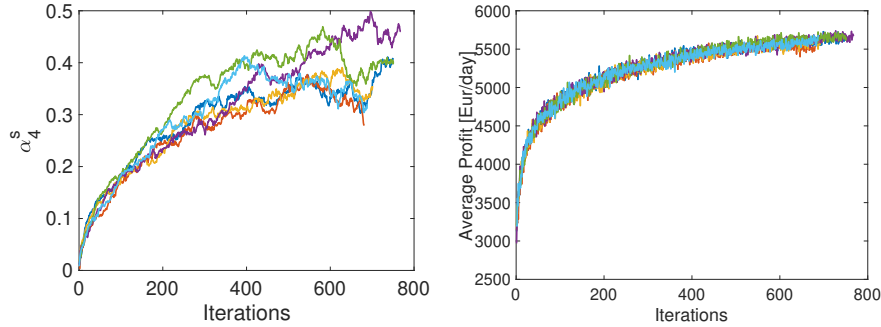


Figure 2.8: Evolution of α_4^s (left) and the average profit (right) for 6 realizations of the REINFORCE algorithm.

2.6.3 Out-of-sample Testing

In this section, we present the results obtained by our threshold policy on out-of-sample data. More precisely, we apply the θ parameter vector learned on the 200 first days of 2015 on the remainder of 2015 and to the entire year of 2016. We compare these results with the ones obtained by rolling intrinsic on the same data. We present the results in Table 2.2. (i) Column 1 represents the trading frequency. (ii) Column 2 refers to the method that we use: "Threshold" refers to the method presented in this chapter, "Threshold init" is the early method that we have developed in [BP19a]¹³, "Rolling 4 PM" refers to rolling intrinsic starting at 4 PM, "Rolling 11 PM" refers to rolling intrinsic starting at 11 PM, and "Threshold without α_i " is the policy learned by the REINFORCE algorithm if we fix α_i^s and α_i^b to 0. (iii) Column 3 refers to the round-trip efficiency of the considered

¹³This method is similar to the one presented in the present chapter, although it does not include any information about the prices that are available for the other delivery hours (there is no α_5^s and α_5^b).

storage unit. (iv) Column 4 refers to the data that are used for the test. It can either be in-sample (the 200 first days of 2015) or out-of-sample (the remainder of 2015 and 2016). (v) Column 5 contains the average profit. From this table, 6 main observations can be made. (i) The performance of rolling intrinsic strongly depends on the time at which the algorithm is started. (ii) Our threshold policy outperforms rolling intrinsic (iii) the threshold policy presented in this chapter is more suited for high frequency than the early method we propose in [BP19a]. (iv) Our threshold policy also outperforms rolling intrinsic for a non perfect round-trip efficiency. (v) The results in and out-of-sample are very similar. (vi) The most important parameters are α_3 α_4 and α_5 . In the remainder of this section, we will analyse these 6 observations in more details.

Trading frequency	Method	Efficiency	Used data	Profit mean
1 hour	Threshold	1	out	5374
1 hour	Threshold init	1	out	4776
1 hour	Rolling 11PM	1	out	4591
1 second	Threshold	1	out	6405
1 second	Threshold init	1	out	5186
1 second	Rolling 11PM	1	out	5438
1 second	Rolling 4PM	1	out	4742
1 second	Threshold	0.81	out	3762
1 second	Rolling 11PM	0.81	out	3311
1 second	Threshold	1	in	6605
1 second	Rolling 11PM	1	in	5694
1 hour	Threshold without α_1	1	out	5362
1 hour	Threshold without α_2	1	out	5375
1 hour	Threshold without α_3 and α_4	1	out	4652

Table 2.2: Profit mean [€/day]

Comparison of rolling intrinsic for different starting time We compare the performance of rolling intrinsic when we start it at 11PM (row 6) and when we start it at 4PM (row 7). We observe that the performance is better when we start the algorithm at 11PM rather than at 4PM. The low profit that is observed when we launch the algorithm at 4PM results from the fact that rolling intrinsic is myopic and therefore sells most of its flexibility directly without accounting for potentially better offers that would appear in the future. When we launch rolling intrinsic at 4PM, it is therefore selling its flexibility while very few offers are available on the market (the market has just opened) which is not a good option. Therefore, in the rest of this section, we only consider the results of rolling intrinsic when we launch it at 11PM.

Superiority of the threshold policy compared to rolling intrinsic By observing rows 4 and 6 of the table, we observe that the average profit difference amounts to 17.8%. Moreover, the proposed threshold policy achieves a higher profit in 77.4% of the days. In Fig. 2.9 we present the daily profit difference. The figure demonstrates that the extra profit is a cumulative effect of multiple days of superior performance, as opposed to being the result of a few isolated days in which the threshold policy performed significantly

better.

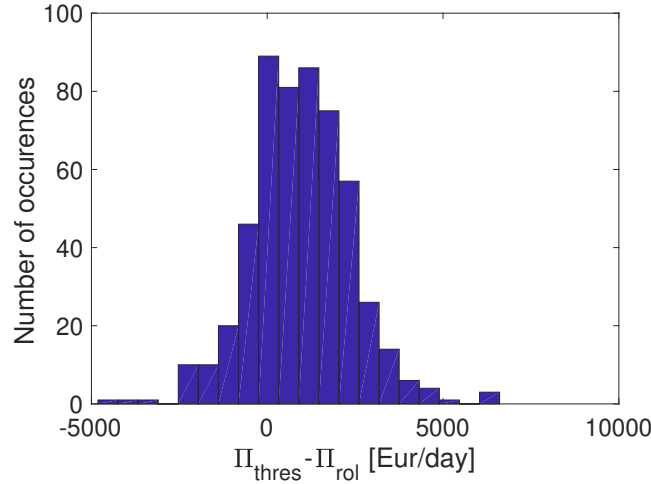


Figure 2.9: Distribution of the difference between the profit of the threshold policy and the rolling intrinsic policy

In Fig. 2.10, we illustrate one of the effects that justifies the superior profit of the threshold policy. Fig. 2.10 indicates whether power has been traded for the different delivery times and time steps. A green dot indicates that we have bought power, whereas a red dot indicates that we have sold power. The left graph illustrates one of the weaknesses of rolling intrinsic: at every line where there are green dots, there are also red dots. This implies that the method only considers buying power if it can sell it directly (except if it can buy power at a negative price). This is due to the fact that the method maximizes the profit of the current time step, and ignores future trading opportunities which may arrive but have not yet been observed. On the contrary, the threshold method procures power at the beginning of the horizon, but may turn down offers for selling power if the sales price is not sufficiently attractive. Thus, the threshold method may wait in order to sell the power later, counting on the possibility that at a later moment there will be offers arriving in the market with a higher willingness to pay than the currently available offers.

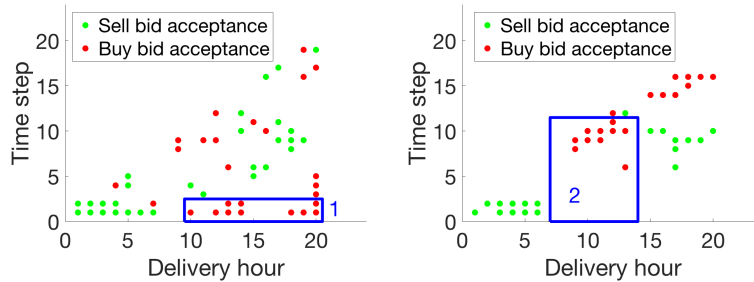


Figure 2.10: Bid acceptance patterns for 1 day of trading for the rolling intrinsic (left) and threshold method (right).

Comparison of the threshold policy and the threshold init policy at high frequency We compare the influence of the trading frequency on the performance of three different methods: (i) the threshold policy presented in the present chapter; (ii) the rolling intrinsic policy; (iii) the threshold init policy. Note that the threshold init policy does not incorporate the parameters α_s^s and α_b^b in the parametrization of the threshold. From rows

1-6 of the table, we observe that the cumulative payoff increases with respect to the trading frequency for all the methods. This is expected, since an increase in the trading frequency increases the number of offers that we observe and use for trading. However, this increase is smaller for the threshold init policy. In order to interpret this result, Figs. 2.11 and 2.12 compare the evolution of five different policies: (i) rolling intrinsic with an hourly trading frequency, (ii) rolling intrinsic with a trading frequency of 15 seconds, (iii) rolling intrinsic with a trading frequency of 1 second, (iv) a threshold policy with a trading frequency of 1 second, and (v) the threshold init policy with a trading frequency of 1 second.

In Fig. 2.11, we observe that two factors contribute to the cumulative payoff. (i) The first factor is the payoff that results from the significant arbitrage possibilities of the storage unit. These arbitrage opportunities can be anticipated. These pay-offs correspond to the large jump of the rolling intrinsic method at 11 PM on D-1. (ii) The second factor corresponds to the payoffs that result from trades of smaller volume, which are not visible at the outset of the trading day. These payoffs correspond to the slight increase of the cumulative payoff of the rolling intrinsic policy, following the large jump. It is worth noting that these small increases are almost insignificant when trading at an hourly time step, but become very important at a higher trading frequency.

This analysis highlights that, when trading at a higher frequency, we require a trading strategy that is effective at capturing the value of both predictable large arbitrage opportunities as well as less predictable small opportunities with a significant cumulative effect. In Fig. 2.12 we observe that the threshold policy attains a similar performance to rolling intrinsic in terms of capturing small arbitrage opportunities. This is represented by the right graph, where we observe that the two curves follow a similar pattern towards the end of the day. On the contrary, the threshold init policy is not able to capture these small arbitrage opportunities, which is clear from the fact that the cumulative pay-off remains constant at the end of the day. Note that the parametrization of the threshold init policy does not include any information about the prices that are available for the other delivery hours. The problem is that, for these small arbitrage opportunities, a bid is not interesting only due to its price but also because if we accept it along with a bid with another delivery time, we can directly secure a positive payoff using our storage unit.

On the other hand, the main difference between our threshold policy and rolling intrinsic mainly rests on the fact that the threshold policy is better suited for trading for big arbitrage opportunities. This is illustrated by the fact that the cumulative payoff reached after the large jump of the threshold policy (the jump of the red curve at 9:20AM) is higher than that of rolling intrinsic. The rolling intrinsic policy buys and sells prematurely in the beginning of the trading period (the jump in the blue curve at 11 PM on D-1), whereas the threshold policy holds back until more favorable trades can be locked in.

Threshold performance for a non perfect roundtrip efficiency In this section, we present the results for a storage unit with a charging efficiency of 0.9 and a discharging efficiency of 0.9. Our aim is to verify that our threshold policy is also suitable for an asset with an imperfect round-trip efficiency. The results are presented in rows 8 and 9 of the table. As before, we compare the results obtained by our threshold policy with the ones obtained by rolling intrinsic on the same data. The proposed threshold policy achieves a higher profit in 64.6% of the days. The average profit difference amounts to 13.6%. In Fig. 2.13 we present the daily profit difference. These results are relatively close to the ones obtained for a storage unit with a perfect round-trip efficiency. The results thus suggest that our policy is also suitable for the case with round-trip efficiency losses.

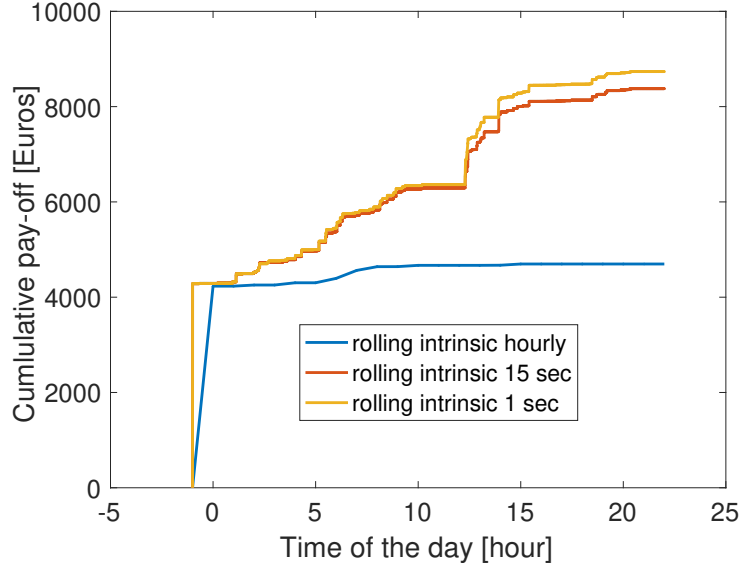


Figure 2.11: Cumulative payoff evolution for one day of trading for the rolling intrinsic method for various trading frequencies.

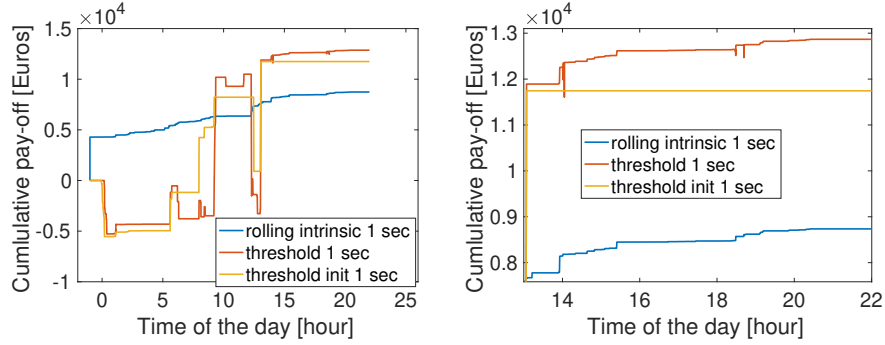


Figure 2.12: Cumulative payoff evolution for one day of trading for the full day (left) and a zoom on the end of the day (right).

Stability of the method with respect to change in the data From rows 4, 6, 10 and 11 of the table, we observe that our threshold policy outperforms rolling intrinsic by 16.0% in-sample and 17.8% out-of-sample. This suggests that our threshold policy performs comparably on unseen data (out-of-sample) as on data on which it has been trained (in-sample). Thus, our method is observed to achieve a robust performance against out-of-sample data.

Importance of the different parameters In order to test the influence of each element of the threshold parametrization, we have launched the REINFORCE algorithm by cancelling each of the parameters one by one. Then we apply the learned policy out-of-sample. More precisely, we have realized 3 simulations: (i) we optimize the policy while fixing α_1^s and α_1^b to 0; (ii) we optimize the policy while fixing α_2^s and α_2^b to 0; (iii) we optimize the policy while fixing α_3^s , α_3^b , α_4^s and α_4^b to 0. The reason for fixing both α_3 and α_4 is that cancelling one will cancel the other automatically, as shown in Eqs. 2.3 and 2.4.

We have not considered the case in which we set α_5^s and α_5^b to 0, because the importance

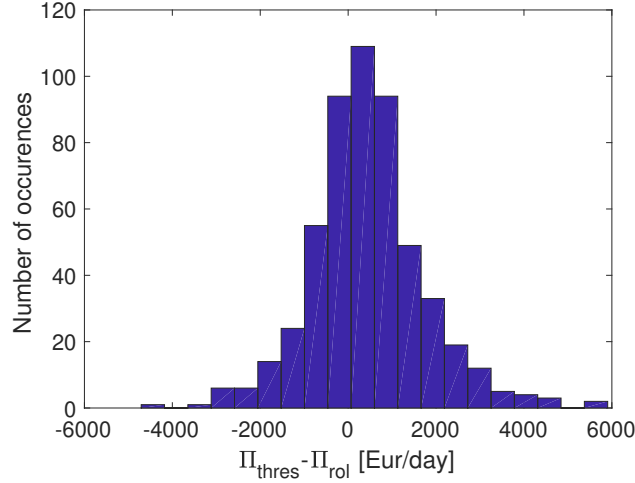


Figure 2.13: Distribution of the difference between the profit of the threshold policy and the rolling intrinsic policy in the case with round-trip efficiency losses.

of this parameter is already discussed extensively when analysing the threshold init policy. The results are presented in the last three rows of the table and will be analysed in detail in the remainder of the section.

1. **Without α_1^s and α_1^b :** We observe that the profit is slightly better if we include α_1^s and α_1^b compared to the case in which we set them to 0. This suggests that these parameters help, but their impact is not decisive¹⁴.
2. **Without α_2^s and α_2^b :** We observe that the profit with and without α_2^s and α_2^b is almost identical. This suggests that this parameter could be discarded without hurting the profitability of the policy.
3. **Without α_3^s , α_3^b , α_4^s and α_4^b :** In this case, we observe a big drop in the profit, compared to the initial case. By further investigating the obtained parameters, we observe that the algorithm converges to a high value of α_1^s and α_1^b . This implies that the sell threshold will be low and the buy threshold will be high. Simultaneously, the algorithm increases the parameters α_5^s and α_5^b to a very high value. The consequence of this behaviour will be that the policy will aim at accepting every possible bid that is also accepted by rolling intrinsic. This indicates that this policy is attempting to mimic the rolling intrinsic policy. In order to confirm this intuition, we present in Fig. 2.14 the profit difference between the rolling intrinsic policy and this policy. We observe that the values are concentrated around 0, which confirms our intuition that the algorithm is attempting to mimic rolling intrinsic.

2.6.4 Profitability of a storage unit trading in the CIM

In this section, we compare the profit earned by our storage unit in the CIM to its investment cost. We consider two types of storage units, a pumped storage hydro unit and a

¹⁴Notice that we have realized the same experiment at a frequency of 1 minute, and the extra profit was more significant (around 0.83%).

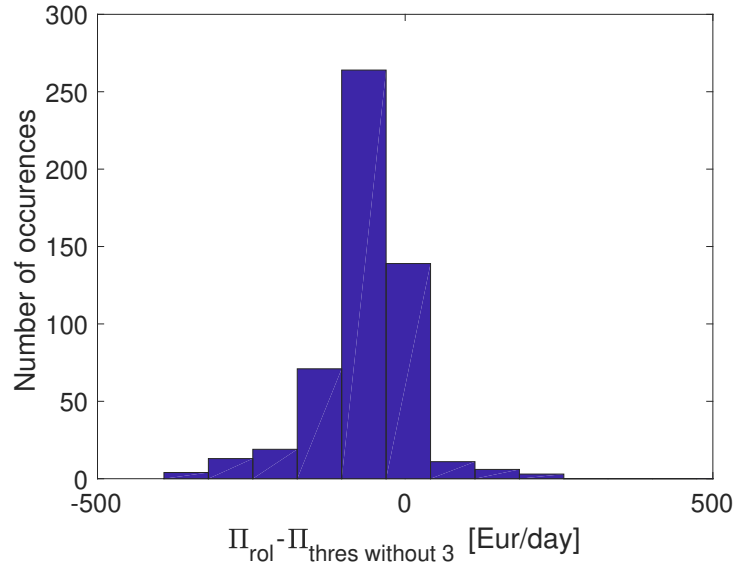


Figure 2.14: Distribution of the difference between the profit of the rolling intrinsic policy and the profit of the threshold policy without parameters $\alpha_3^{s/b}$ and $\alpha_4^{s/b}$

large scale battery. We present the relevant financial data about these two types of assets in Table 2.3. In order to quantify revenues, we use the results presented in Table 2.2. For a pumped storage hydro unit, we select the results with a round-trip efficiency of 81%; and for a battery, we select the results with a perfect round-trip efficiency. Using these data¹⁵, we can compute the return on investment for the different projects.

Pumped storage hydro: For a pumped storage hydro, if we use the most optimistic parameters, we obtain a return of 2.5% while if we use the least optimistic parameters, we obtain a return of -5.4% . This indicates that the CIM might be one of the main sources of revenue for a pumped storage hydro unit. These results are encouraging, because they do not account for all of the profits that a pumped storage hydro unit can have access to. Indeed, we only trade products with an hourly granularity, and could still extract an extra profit from adapting our output on a 15-minute basis. Moreover, market participants coordinate their strategy in several markets. Therefore, not all of the profit is expected to originate from a single market. For instance, part of the flexibility could be sold in ancillary services, or could be used to balance the portfolio of the market participant in real time.

Large-scale battery: For the battery, the yearly profit does not cover the annual fixed cost. This result indicates that the CIM should not be the main source of revenue for a battery. This is confirmed by [IRE21], where it is shown that most installed large-scale batteries are used for providing (primary/secondary) reserve. Nevertheless, this does not imply that other markets than reserve should not be considered. Given that the capacity

¹⁵Notice that we use the data of the CIM in 2015–2016. Since 2016, the liquidity in the CIM has increased, and cross-border bids have been introduced. This increases trading opportunities, and may therefore increase the return on investment if the strategy would be applied against current market conditions.

cleared in day-ahead reserve markets cannot be traded in the CIM, an effective strategy in the CIM mitigates the risk of not being accepted in the reserve markets, and therefore allows a more aggressive bidding strategy in reserve markets (bidding at a higher price while covering the risk of not being accepted by trading the flexibility in the CIM).

Characteristics/Unit type	Pumped storage hydro	Large scale battery
Investment cost [k€/MW]	106 [US 21a]-800 [RAE21]	200 [ZT20]-700 [RAE21]
Round-trip efficiency [%]	79 [US 21b]	90 [Tes21]
Annual fixed cost [k€/MW]	4 [RAE21]	14 [RAE21]
Economic Lifetime [years]	50 [RAE21]-100 [NRE03]	10 [RAE21]

Table 2.3: Financial and technical informations on storage units.

2.7 Conclusions and Perspectives

In this chapter, we tackle the problem of intraday trading for storage units. We model the problem using Markov Decision Processes. We focus on policies that are parametrized on price thresholds, and we optimize the resulting policy functions using the REINFORCE algorithm. We introduce and justify a collection of factors that can be used for adapting the trading threshold to system conditions. We compare our threshold policy to the rolling intrinsic method on the German continuous intraday market. We demonstrate that the threshold policy performs significantly better than rolling intrinsic, and analyze the results in order to explain the performance difference. Future extensions of the work can include the following items. (i) One can improve the policy functions by adding more explanatory variables of the price thresholds such as renewable forecasts or generator outages. (ii) One can compute the value of coordinated trading strategies between early markets (day-ahead and intraday auction, day-ahead reserve markets) and the continuous intraday market. Day-ahead reserve markets are likely good candidates because they are also a notable revenue stream for flexible assets such as storage units. (iii) One could consider the possibility of placing bids in the market. This might be interesting because, as explained in section 2.2, it may lead to a significant increase in profits. A first approach that would allow us to still rely on our threshold policy would be to consider that all the bids that we have placed during previous time steps are always cancelled, and to place bids for a predefined quantity.

2.8 Appendix

2.8.1 Appendix A: Quantity that we exchange compared to the total exchanged quantity

In order to better understand the potential for our trading strategy to influence the strategy of other agents, we have compared the quantity that we exchange for different delivery hours with the total exchanged quantity in the market. More precisely, we have recorded the total exchanged quantity in the German CIM for every hour of 10 days that are randomly selected in our dataset. We have then computed the ratio between our traded quantity and the total exchanged quantity for these hours. We present the results graphically in Fig. 2.15. It can be observed that, for most hours, the ratio is quite low (the average ratio is around 1.65%). For these hours, it could be argued that our influence on

the future behavior of other actors could be negligible. On the contrary, for a few hours, our trading volume can reach a ratio of 10%.

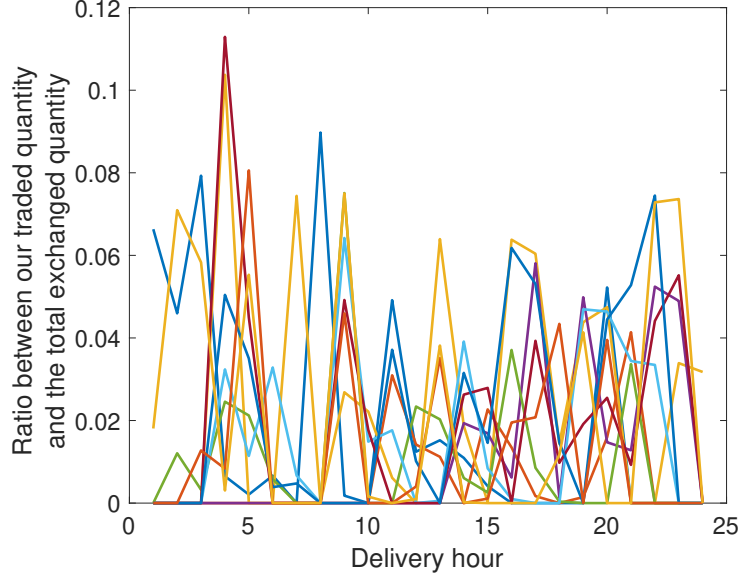


Figure 2.15: Ratio between our traded quantity and the total exchanged quantity for 10 randomly selected days.

2.8.2 Appendix B: Characterization of Optimal Trading Policy in a Deterministic Setting

In this appendix, we prove that optimal inter-temporal arbitrage in a deterministic setting is achieved by a threshold policy if the reservoir limit constraint is not binding. Towards this aim, we characterize the optimal policy in a perfect foresight setting¹⁶. This characterization is obtained from an analysis of the KKT conditions of the perfect foresight model.

2.8.2.1 Notations

In order to describe the model, we need to define the sets, the parameters and the variables.

- The sets are:
 - T , the set of time steps.
 - D , the set of the delivery times.
 - $I_{t,d}, \forall t \in T, d \in D$, the set of bids available for a given time step and delivery time.
- The parameters are:
 - $p_{i,t,d}^b$, the price of the i^{th} buy bid for delivery time d at time step t .
 - $p_{i,t,d}^s$, the price of the i^{th} sell bid for delivery time d at time step t .

¹⁶Perfect foresight should be understood as the situation in which we know in advance all the bids which will arrive in the future.

- $Q_{i,t,d}^b$, the available quantity of the i^{th} buy bid for delivery time d at time step t .
- $Q_{i,t,d}^s$, the available quantity of the i^{th} sell bid for delivery time d at time step t .
- V , the maximum capacity of the reservoir.
- The variables are:
 - $q_{i,t,d}^b$, the quantity we accept from the i^{th} buy bid for delivery time d at time step t .
 - $q_{i,t,d}^s$, the quantity we accept from the i^{th} sell bid for delivery time d at time step t .
 - v_d , the capacity stored in the reservoir at delivery hour d .

2.8.2.2 Model

The model is given by (P) .

$$(P) : \max \sum_{t \in T} \sum_{d \in D} \sum_{i \in I_{t,d}} \left(p_{i,t,d}^b q_{i,t,d}^b - p_{i,t,d}^s q_{i,t,d}^s \right) \quad (2.13)$$

$$q_{i,t,d}^{s/b} \leq Q_{i,t,d}^{s/b}, \forall i \in I_{t,d}, d \in D, t \in T \quad (\mu_{i,t,d}^{s/b}) \quad (2.14)$$

$$q_{i,t,d}^{s/b} \geq 0, \forall i \in I_{t,d}, d \in D, t \in T \quad (\nu_{i,t,d}^{s/b}) \quad (2.15)$$

$$v_d = v_{d-1} + \sum_{t \in T} \sum_{i \in I_{t,d}} \left(q_{i,t,d}^s - q_{i,t,d}^b \right), \forall d \in D \quad (\lambda_d) \quad (2.16)$$

$$0 \leq v_d, \forall d \in D \quad (\beta_d) \quad (2.17)$$

$$v_d \leq V, \forall d \in D \quad (\gamma_d) \quad (2.18)$$

Objective function (2.13) determines the profit as the sum of the revenue for what we sell minus the cost of what we buy. Constraints (2.14) and (2.15) state that we can only accept a positive quantity of the bids and less than the maximal quantity of the bid. Constraint (2.16) models the evolution of the capacity stored in the reservoir. Constraints (2.17) and (2.18) ensure that the stored energy remains in the reservoir limit.

2.8.2.3 KKT Conditions

In order to compute the KKT conditions, we compute the Lagrangian:

$$\begin{aligned} L = & \sum_{t \in T} \sum_{d \in D} \sum_{i \in I_{t,d}} \left(p_{i,t,d}^b q_{i,t,d}^b - p_{i,t,d}^s q_{i,t,d}^s + \mu_{i,t,d}^s (Q_{i,t,d}^s - q_{i,t,d}^s) + \mu_{i,t,d}^b (Q_{i,t,d}^b - q_{i,t,d}^b) \right. \\ & \left. + \nu_{i,t,d}^s q_{i,t,d}^s + \nu_{i,t,d}^b q_{i,t,d}^b \right) - \sum_{d \in D} \lambda_d \left(v_d - v_{d-1} - \sum_{t \in T} \sum_{i \in I_{t,d}} q_{i,t,d}^s + \sum_{t \in T} \sum_{i \in I_{t,d}} q_{i,t,d}^b \right) \\ & + \sum_{d \in D} \left(\beta_d v_d + \gamma_d (V - v_d) \right) \end{aligned}$$

The constraint that the gradient of the Lagrangian vanishes can be written as:

$$0 = p_{i,t,d}^b - \mu_{i,t,d}^b + \nu_{i,t,d}^b - \lambda_d \quad (q_{i,t,d}^b) \quad (2.19)$$

$$0 = -p_{i,t,d}^s - \mu_{i,t,d}^s + \nu_{i,t,d}^s + \lambda_d \quad (q_{i,t,d}^s) \quad (2.20)$$

$$0 = \beta_D - \gamma_D - \lambda_D \quad (v_D) \quad (2.21)$$

$$0 = \beta_d - \gamma_d - \lambda_d + \lambda_{d+1} \quad (v_d) \quad (2.22)$$

From equations (2.19), we obtain

- If $p_{i,t,d}^b > \lambda_d$ then $\nu_{i,t,d}^b = 0$ and $\mu_{i,t,d}^b > 0$, which implies that $q_{i,t,d}^b = Q_{i,t,d}^b$.
- If $p_{i,t,d}^b < \lambda_d$ then $\mu_{i,t,d}^b = 0$ and $\nu_{i,t,d}^b > 0$, which implies that $q_{i,t,d}^b = 0$.
- If $p_{i,t,d}^b = \lambda_d$, then the bid is partially accepted.

From equations (2.20) we obtain

- If $p_{i,t,d}^s < \lambda_d$ then $\nu_{i,t,d}^s = 0$ and $\mu_{i,t,d}^s > 0$, which implies that $q_{i,t,d}^s = Q_{i,t,d}^s$.
- If $p_{i,t,d}^s > \lambda_d$ then $\mu_{i,t,d}^s = 0$ and $\nu_{i,t,d}^s > 0$, which implies that $q_{i,t,d}^s = 0$.
- If $p_{i,t,d}^s = \lambda_d$, then the bid is partially accepted.

From these expressions, we can interpret λ_d as the trading threshold for delivery time d . Indeed, we have proven that we accept any sell bids with a lower price and any buy bids with a higher price. After that, we can analyse the link between the λ for different delivery times. We start by analysing the value of the threshold at the last delivery time using equation (2.21), where we note 3 possibilities:

- If $v_D = 0$ then $\gamma_D = 0$, which implies that $\lambda_D \geq 0$.
- If $v_D = V$ then $\beta_D = 0$, which implies that $\lambda_D \leq 0$.
- If $0 < v_D < V$ then $\beta_D = 0, \gamma_D = 0$, which implies that $\lambda_D = 0$.

Proceeding backwards using equation (2.22), at delivery time d we have the three following possibilities:

- If $v_d = 0$ then $\gamma_d = 0$, which implies that $\lambda_{d+1} \leq \lambda_d$.
- If $v_d = V$ then $\beta_d = 0$, which implies that $\lambda_{d+1} \geq \lambda_d$.
- If $0 < v_d < V$ then $\beta_d = 0, \gamma_d = 0$, which implies that $\lambda_{d+1} = \lambda_d$.

From this, we conclude that, if the reservoir is not binding, the λ_d are equal to each other. This means that the optimal policy is a threshold policy if the reservoir is not binding, which is the result that we wish to prove.

2.8.3 Appendix C: Rolling Intrinsic Model

In this appendix, we present the optimization model that the rolling intrinsic method solves at each time step $t \in T$ in the case of a perfect round-trip efficiency storage. We start by defining the notation. Then we present the model.

2.8.3.1 Notations

- The sets are:
 - D , the set of the delivery times.
 - $I_{t,d}, \forall d \in D_t$, the set of bids available for a given delivery time at time step t .
- The parameters are:
 - $p_{i,t,d}^b$, the price of the i^{th} buy bid for delivery time d at time step t .
 - $p_{i,t,d}^s$, the price of the i^{th} sell bid for delivery time d at time step t .
 - $Q_{i,t,d}^b$, the available quantity of the i^{th} buy bid for delivery time d at time step t .
 - $Q_{i,t,d}^s$, the available quantity of the i^{th} sell bid for delivery time d at time step t .
 - V , the maximum capacity of the reservoir.
 - $v_{t-1,d}$, the capacity that we would get at delivery time d , with the trade we have realized until time step $t - 1$.
- The variables are:
 - $q_{i,t,d}^b$, the quantity we accept from the i^{th} buy bid for delivery time d at time step t .
 - $q_{i,t,d}^s$, the quantity we accept from the i^{th} sell bid for delivery time d at time step t .
 - $v_{t,d}$, the capacity that we would get at delivery time d , if we do not trade after time t .

2.8.3.2 Model

The interpretation is that we accept, at each time step t , the set of bids that maximize the instantaneous profit. It means that, $\forall t \in T$ we solve (P_t) . Objective (2.23) determines the profit as the sum of the revenue for what we sell minus the cost of what we buy. Constraint (2.24) requires that we cannot accept a quantity greater than the maximum capacity of the bid. Constraint (2.25) imposes that we accept a positive quantity of the bid. Constraint (2.26) shows how the capacity of the reservoir evolves with respect to the quantity that we trade. Constraints (2.27) and (2.28) require that the capacity stored in the reservoir should stay in the reservoir limit.

$$(P_t) \quad \max_{q_{i,t,d}^{s/b}, v_{t,d}} \sum_{d \in D} \sum_{i \in I_{t,d}} \left(p_{i,t,d}^b q_{i,t,d}^b - p_{i,t,d}^s q_{i,t,d}^s \right) \quad (2.23)$$

$$q_{i,t,d}^{s/b} \leq Q_{i,t,d}^{s/b} \quad \forall i \in I_{t,d}, d \in D \quad (2.24)$$

$$q_{i,t,d}^{s/b} \geq 0 \quad \forall i \in I_{t,d}, d \in D \quad (2.25)$$

$$v_{t,d} = v_{t-1,d} + \sum_{a \in D | a \leq d} \sum_{i \in I_{t,a}} \left(q_{i,t,a}^s - q_{i,t,a}^b \right) \quad \forall d \in D \quad (2.26)$$

$$0 \leq v_{t,d} \quad \forall d \in D \quad (2.27)$$

$$v_{t,d} \leq V \quad \forall d \in D \quad (2.28)$$

2.8.4 Appendix D: Order book simulator

We commence by describing the format of the bids received from EPEX. We then present the 4 steps that we apply in order to obtain the data for our simulator. These 4 steps include (i) linking the partially accepted bids, (ii) linking the iceberg bids, (iii) suppressing invisible bids, and (iv) cancelling arbitrage opportunities. Finally, we convert the data to the format that is compatible with the simulator that is presented in section 2.2.

Initial format of the data: The format of the data is illustrated in table 2.4. Each line of the table represents a bid placed in the market. The bids have the following characteristics: (i) **Instrument type**, the length of the delivery period. It can be 1 hour or 15 minutes. (ii) **Delivery instrument** (delivery period), the moment at which the energy should be produced. (iii) **Delivery date**, the date at which the power needs to be produced (iv) **Start validity date**, the moment at which the bid appears in the platform (v) **End validity date**, the moment at which the bid is not available anymore (vi) **Cancelling date**, the moment at which the bid has been cancelled (only if the bid has been withdrawn by the trader who has placed it, not if the bid has been matched by another trader). (vii) **Is executed**, this indicator is equal to 0 if the bid has not been matched, 1 if the bid has been matched for its total quantity, and 2 if the bid has been matched for a partial quantity. (viii) **Side**, S if it is a sell bid and B if it is a buy bid. (ix) **Price**, the price at which the bid has been placed. (x) **Execution price**, the price at which the bid has been executed. (xi) **Volume**, the volume of the bid. (xii) **Executed volume**, the volume of the bid which has been matched. After presenting the way in which the data is made available, we describe in the next sections what adaptations we perform to the data in order to be able to simulate the market.

Partially accepted bids: We illustrate in table 2.5 the case of a bid that is partially accepted for 10MW in a first exchange and that is matched for the remaining quantity in a second exchange. The way that EPEX handles this is as follows. When the bid is partially matched in the first exchange (between the bids of lines 1 and 2), the bid is cancelled from the order book (line 1) and a new bid is created with the same characteristics and with a quantity equal to the unmatched quantity of the initial bid (line 3). This way of representing the data is not suitable for our problem, because, if we would decide to accept (partially or fully) the bid of the first line before the arrival of the bid of line 2, our simulator would still observe the arrival of the bid of line 3. The problem is that the bid in line 3 is dependent on the bid of line 1, however this dependence is not represented in the data. In order to overcome this problem, we filter all of the bids in order to find the ones that are linked together (bids that are created following the partial acceptance of a previous bid). For this purpose, we iterate over all of the bids and consider a bid as being linked to a previous one if (i) the new bid arrives at the same time as the previous one disappears, (ii) the volume of the new bid is equal to the initial volume of the previous bid minus the volume accepted from the previous bid, (iii) the two bids have the same price, (iv) the two bids have the same side (buy/sell), (v) the new bid is not yet part of another group, and (vi) the accepted quantity of the first bid is non-zero.

Iceberg bids: An iceberg bid is a particular type of bid in the CIM. It is a bid for which only a partial volume of the bid is observed on the platform. For instance, a trader can split a bid of 200MWh into 4 bids of 50MWh. In this case, only a bid of 50MWh will

be visible on the platform. When the 50MWh are accepted, the next 50MWh block is made available on the platform. The interest for these products is that the traders do not wish to reveal that they are willing to buy/sell a large quantity of power. An example of how an iceberg bid is represented in EPEX data is provided in table 2.6. In this example, an iceberg bid for selling 90MWh is split into three 30MWh bids (line 1, line 3, line 5 and line 7). This iceberg bid is first matched with a 30MWh bid (line 2) and later with a 40MWh bid (line 4 and 6). The remaining 20MWh are later withdrawn by the trader (line 7). As for the partial acceptance of a bid, there is no link between the different parts of an iceberg bid in the EPEX data. In order to overcome this problem, we filter all of the bids in order to find which bids are part of an iceberg bid. An interesting property for detecting if a bid is part of an iceberg bid is illustrated in Table 2.6. For a bid that is part of an iceberg bid, if its executed volume is equal to its volume, its *is executed* property is set as being partially accepted (in line 1 and 3, *is executed* is equal to 2). For linking bids that are part of the same iceberg bids, we iterate through all of the bids and consider a bid as being linked to a previous one if (i) the executed volume of the previous bid is equal to its total volume, (ii) the *is executed* of the previous bid is equal to 2, (iii) the price of the new bid is the same as the price of the previous bid, (iv) the side of the new bid is the same as the side of the previous bid, (v) the new bid is not yet part of any iceberg bid, (vi) the new bid arrives at the same time as the previous bid disappears, (vii) the new bid is a starting bid (not a bid that is created because a previous bid has been partially accepted).

Suppression of invisible bids: The concept of an invisible bid is illustrated in Table 2.5. By observing the data, we may think that the bid of line 2 is available during 14 millisecond. In fact, this is not true because the bid is directly matched with the bid of line 1. The 14 milliseconds correspond to the time for the platform to realize that the bid of line 1 and 2 should be matched. Therefore, it would not make sense that our algorithm has the chance to accept the bid of line 2 during these 14 milliseconds. Thus, we suppress offers that appear and (i) have a price to sell (resp. buy) power lower (resp. higher) than the available bid with the best price to buy (resp. sell) power, and (ii) are available for less than 300 milliseconds. The reason for suppressing only bids that remain for less than 300 milliseconds is that it can be legitimate to have, for the same delivery period, a sell bid that has a lower price than a buy bid if one of them is part of a block bid (bid spanning multiple delivery periods) or is an all-or-none bid (bids that can only be fully accepted). Therefore, if we would not use this threshold of 300 milliseconds, we could have an initial all-or-nothing sell (resp. buy) bid that stays long in the market, which would mean that we would cancel many buy (resp. sell) bids (with a higher (resp. lower) price) even though they are perfectly valid on the market.

Cancellation of arbitrage opportunities As explained above, it is possible to have, for the same delivery period, a sell bid that has a lower price than a buy bid if one of them is part of a block bid (bid spanning multiple delivery periods) or is an all-or-none bid (bids that can only be fully accepted). Nevertheless, as our simulator only considers continuous bids and iceberg bids (we do not have information about block bids or all-or-none bids), this would lead to a direct arbitrage opportunity. Indeed, we would be able to sell and buy power for the same delivery hour and have a positive payoff without assuming any risk. In order to cancel these arbitrage opportunities, we use the following algorithm. We simulate the bid arrival in the market and check if there is an arbitrage opportunity. If

it is the case, we compute for the buy and the sell side how many bids we would need to remove in order to cancel the arbitrage opportunity. Finally, we cancel the bids from the side that requires removing the fewest bids.

Computation of Events Following the data treatment presented earlier, we can input the data in the format of the simulator presented in section 2.2. To this aim, we iterate through each bid group (partial activation or iceberg) and we define (i) an **open** event when the first bid of the group arrives, the associated volume is the total volume of the group; (ii) an **acceptance** event, every time a sub-bid is posted on the market; (iii) a **close** event when the last bid of the group disappears. Notice that, in the initial data, it is possible that two bids arrive at the same time. If it is the case, there are two possibilities. (i) The bids are linked together. This situation only occurs in the three cases explained above (partial acceptance, iceberg bids, suppression of invisible bids). (ii) The different bids are independent of each other. In this case, we can simply execute both events in any order.

Instrument Type	Delivery Instrument	Delivery date	Start validity date	End validity date	Cancelling date	is Executed	Side	Price	Execution price	Volume	Executed volume
Hour	6:00	12/01/15	11/01/15 22:43:27.007	12/01/15 02:53:23.432	12/01/15 02:53:23.432	0	B		15	50	0
Hour	13:00	12/01/15	11/01/15 05:01:35.007	11/01/15 10:01:35.123		1	S		25	50	50
Hour	18:00	12/01/15	11/01/15 07:23:43.145	11/01/15 13:31:57.364		2	S	32	32	50	9.8

Table 2.4: Format of the data provided by EPEX.

Instrument Type	Delivery Instrument	Delivery date	Start validity date	End validity date	Cancelling date	is Executed	Side	Price	Execution price	Volume	Executed volume
Hour	18:00	12/01/15	11/01/15 07:23:43.145	11/01/15 13:31:57.364		2	S	32	32	50	10
Hour	18:00	12/01/15	11/01/15 13:31:57.350	11/01/15 13:31:57.364		1	B	32	32	10	10
Hour	18:00	12/01/15	11/01/15 13:31:57.364	11/01/15 14:42:57.495		1	S	32	32	40	40

Table 2.5: Example of the partial acceptance of a bid and of unavailable bids in EPEX data

Instrument Type	Delivery Instrument	Delivery date	Start validity date	End validity date	Cancelling date	is Executed	Side	Price	Execution price	Volume	Executed volume
Hour	22:00	12/01/15	11/01/15 10:00:00.000	11/01/15 11:00:00.142		2	S	40	40	30	30
Hour	22:00	12/01/15	11/01/15 11:00:00.000	11/01/15 11:00:00.142		1	B	40	40	30	30
Hour	22:00	12/01/15	11/01/15 11:00:00.142	11/01/15 12:00:00.038		2	S	40	40	30	30
Hour	22:00	12/01/15	11/01/15 12:00:00.000	11/01/15 12:00:00.038		2	B	40	40	40	30
Hour	22:00	12/01/15	11/01/15 12:00:00.038	11/01/15 12:00:00.038		2	S	40	40	30	10
Hour	22:00	12/01/15	11/01/15 12:00:00.038	11/01/15 12:00:00.038		1	B	40	40	10	10
Hour	22:00	12/01/15	11/01/15 12:00:00.038	11/01/15 15:00:00.000	11/01/15 16:00:00.000	0	S	40		20	0

Table 2.6: Example of an iceberg bid in EPEX data

2.8.5 Determination of Regimes

In this section, we start by presenting why we compute the regimes based on the intraday auction price curve. Then, we describe how we compute these regimes in details.

2.8.5.1 Regimes based on the intraday auction

As explained in section 2.5.1, the regimes are computed from the pattern of the intraday auction. There are three reasons for this choice (i) this assumption seems coherent with the observations from the data. In Fig. 2.16, we compare the price in the CIM one hour before delivery, with the price of the intraday auction and the price in the CIM at 11PM. It can be observed that the pattern of the intraday auction price is very similar to the one of the CIM price at 11PM and fits relatively well with the CIM price one hour before delivery (ii) The gain of using the CIM price pattern at 11PM rather than the intraday auction price to predict the pattern of the CIM price one hour before delivery seems relatively limited (in Fig. 2.16, the yellow curve is not really more resembling to the blue one than the red one). (iii) As the price pattern can change at any moment of the CIM, it would be too heavy computationally to recompute the regimes every time an offer appears/disappears from the CIM. Indeed, this would require running the procedure of section 2.8.5.2 which includes running an (easy) optimization problem.

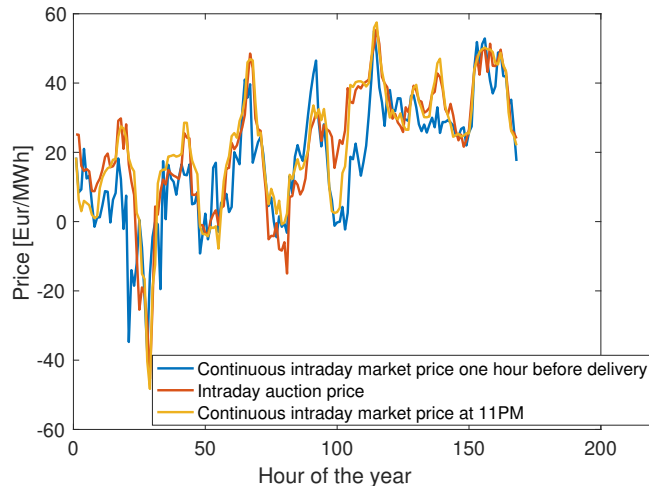


Figure 2.16: Price in the CIM and in the intraday auction for the first week of 2015.

2.8.5.2 Stable optimum

In order to define our regimes, we do not consider every maxima and minima of the intraday auction price curve. We only consider the one that can bring a sufficient arbitrage profit meaning that the difference between the regime maximum and the regime minimum is large enough. The reasons for this choice are (i) to avoid having too many regimes (ii) only have minima and maxima that remains valid through the CIM trading time (as shown in section 2.8.5.1, the pattern in the CIM might slightly change through time).

In order to detect the stable optima, we use the following optimization model ((2.29)-(2.34)) where: (i) q_d is the quantity we sell for delivery hour d ; (ii) v_d is the capacity stored in the reservoir at delivery time d ; (iii) s_d is the quantity traded for delivery hour d (iv) p_d is the price for delivery hour d ; (v) c is the selectivity parameter. It is the minimum profit

per MWh traded that is needed in order to define a new regime; and (vi) V is the maximum capacity of the reservoir. Objective (2.29) determines the profit as the sum of the revenue for what we trade in the intraday auction at which we subtract the minimum profit we want per MWh traded. Constraints (2.30) and (2.31) require that s_d is the quantity traded for delivery hour d (absolute value of q_d). Constraint (2.32) shows how the capacity of the reservoir evolves with respect to the quantity that we trade. Constraints (2.33) and (2.34) require that the capacity stored in the reservoir remains in the reservoir limit. Solving this optimization problem, it is possible to not obtain any regime. In that situation, we use an iterative procedure that decreases the selectivity parameter c until we obtain at least one regime. The selectivity parameter is arbitrarily fixed at $4 \frac{\text{€}}{\text{MWh}}$.

$$\max_{q_d, v_d, s_d} \sum_{d \in D} (p_d \cdot q_d - c \cdot s_d) \quad (2.29)$$

$$s_d \geq q_d \quad \forall d \in D \quad (2.30)$$

$$s_d \geq -q_d \quad \forall d \in D \quad (2.31)$$

$$v_d = v_{d-1} - q_d \quad \forall d \in D \quad (2.32)$$

$$0 \leq v_d \quad \forall d \in D \quad (2.33)$$

$$v_d \leq V \quad \forall d \in D \quad (2.34)$$

Chapter 3

Optimal Trading of a Fixed Quantity of Power in an Illiquid Continuous Intraday Market

Renewable assets face considerable supply uncertainty, and therefore stand to gain by adjusting their position dynamically in the CIM, as more accurate forecast information arrives for their real-time supply. Moreover, trading later in the day also increases opportunities for profitable trades, since bid-ask spreads in CIMs are empirically observed to decrease as we approach real time. These benefits need to be traded off against the fact that the CIM is typically less liquid than earlier forward (e.g. day-ahead) markets. Therefore, there is a counter-balancing interest for a renewable supplier to sell its power earlier, in order to avoid “pushing” the price against its profits by unloading large quantities of supply in a thin market.

In this work we set the foundation for capturing the latter tradeoff (thin markets), and leave matters associated to the uncertainty of supply and the increasing information that is revealed closer to real time for future work, but set in place the algorithmic framework for this extension. Concretely, we focus on developing an optimal trading strategy for selling a fixed quantity of power in an idealized CIM for which we have a model of the price evolution. Our motivation is to set the basis for value function approximation (VFA) algorithms that can be used for trading the production of a renewable unit in the real CIM without any assumption on the price evolution model.

Our analytical work draws similarities to early work on optimal control by [Mor59]. In this work, the authors develop an optimal trading strategy in order to trade a fixed quantity within a certain deadline subject to independent random prices. They prove that the optimal strategy is characterized by a threshold beyond which the producer should trade the required quantity. This work has been extended by [Kin69] and [Gol85], where the authors consider that the trader (i) has the option to store the good for a given holding cost, and (ii) faces a deterministic demand at each time period. They prove that the optimal policy still follows a threshold strategy.

In more recent work in the context of electricity markets [RAP16], the authors derive an optimal strategy for a thermal unit trading in the CIM, while assuming that the price follows an additive Brownian motion. In [CJP15], the authors also present the solution for trading a fixed quantity. The difference with our work is that [RAP16] does not account for any bid-ask spread and [CJP15] only considers the case of a constant bid-ask spread.

On the contrary, in our work, we solve the problem using a non constant spread in order to reflect the empirical observation that the CIM remains relatively illiquid, despite its growth.

In the previous chapter, we have motivated the idea for developing a threshold policy by the fact that threshold policies have been *proven* to be optimal in a number of papers in the literature regarding specific instances of stochastic optimal control problems with uncertain prices [Mor59, Kin69, Gol85]. By analogy, in the present chapter, we develop a value function approximation for the context of a renewable supplier with uncertain supply.

Our contributions can be summarized as follows: (i) We cast the problem of unloading a fixed quantity of power in a simplified CIM, as an MDP. (ii) We characterize the optimal policy as well as the optimal value function for this MDP. (iii) We use this optimal value function to develop basis functions for a VFA algorithm. (iv) We use the MDP framework to also develop an SDDP algorithm, that can be used as a benchmark for the VFA algorithm. We validate our VFA and SDDP algorithms by demonstrating that they both arrive to the optimal analytical solution of a 10-period example.

3.1 Continuous Intraday Market Model

As explained in section 1.3.3, at any moment, a trader in the CIM observes a collection of bids. This collection of bids is called an order book. This order book can be further split into 24 “hourly” order books¹, one for each delivery hour. We present such an order book in Fig. 3.1. The order book consists of two parts: (i) the buy side, which contains all the bids of traders which want to buy power from us (in blue), and (ii) the sell side, which contains all the bids of traders which want to sell power to us (in red).

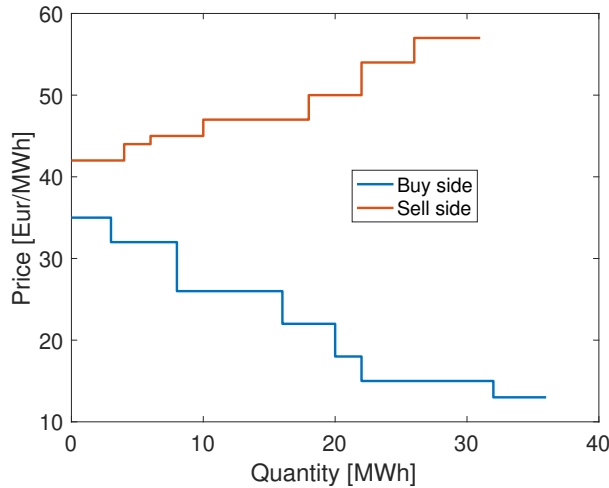


Figure 3.1: Example of an order book.

In order to derive an optimal trading strategy, we propose decomposing the order book into three components that are presented in Fig. 3.2: (i) a bid-ask spread, which is the price difference between the most favorable sell and buy bids that have yet to be matched (illustrated in the left panel); (ii) the center of the bid-ask spread, which is the average price between the best sell bid and the best buy bid (illustrated in the left panel); and

¹In the remainder of the chapter, we only consider “hourly” order books.

(iii) a linear price impact (illustrated in the right panel). We exploit the linearity in the derivation of the optimal trading policy.

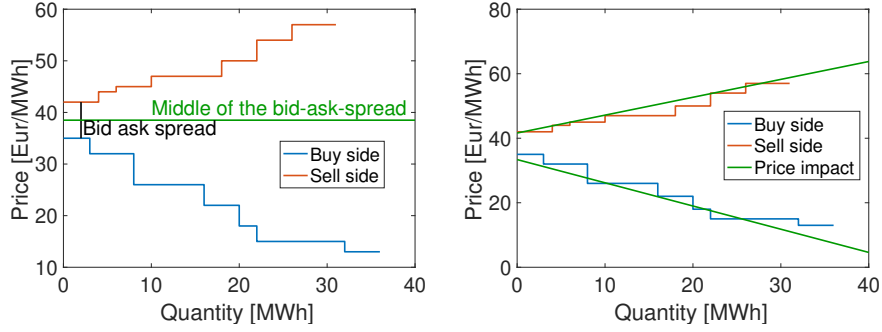


Figure 3.2: Illustration of the different components of the order book.

This decomposition of the order book at time step t can also be expressed mathematically as:

$$p_t^s(q_t) = p_t - \Delta_t - 2rq_t \quad (3.1)$$

where

- $p_t^s(q_t)$ is the marginal price at which we sell quantity q_t ;
- p_t is the center of the bid ask spread. We assume that p_t follows a stochastic evolution, according to the following model²:

$$p_{t-1} = p_t + \epsilon_t \quad (3.2)$$

where ϵ_t can follow any distribution respecting $\mathbb{E}[\epsilon_t] = 0$, where the expectation is conditional on the information available in time t .

- $2r$ is the slope of the linear impact and q_t is the quantity we sell. The $2rq_t$ term therefore represents the impact of the producer on the price. This term represents the fact that the price lowers as a producer sells more power in the CIM. The parameter r is assumed to be deterministic. In order to estimate it, we use confidential data from the German CIM for 2015 – 2016 which has been sourced from the European Power Exchange (EPEX). We use the following strategy, which is initially proposed in [CJP15]. We record the state of the market in different instances. For each of these instances i , we compute the marginal price $p^{si}(q^j)$ that we would obtain if we were selling different quantities of power q^j . For each of these quantities, we compute the average price obtained from the different instances:

$$\bar{p}^s(q^j) = \sum_{i=1}^I \frac{p^{si}(q^j)}{I},$$

where I is the number of instances. These averages are presented in the left panel of Fig. 3.3. Finally, we use a linear regression to obtain the red line, the slope of which is equal to the price impact coefficient $2r$. Estimating the value against the 200 first days of 2015 yields a value of $r = 0.0095$.

- Δ_t is half the bid-ask spread which is assumed to be deterministic. As in the case of r , we record the state of the market in different instances, although we now separate

²In order to simplify the notations for the backward induction proof, we reverse the time index. Specifically, we consider that the index of the last time step is 0 and the index of the first time step is T .

the instances in different batches, depending on how much time before delivery the instance has been captured. For each of these batches, we compute the average spread. The results of this computation are presented in the right panel of Fig. 3.3. One important insight from this graph is that, as we arrive closer to delivery, the bid-ask-spread decreases. This indicates that it can be favorable to wait for market closure to trade power.

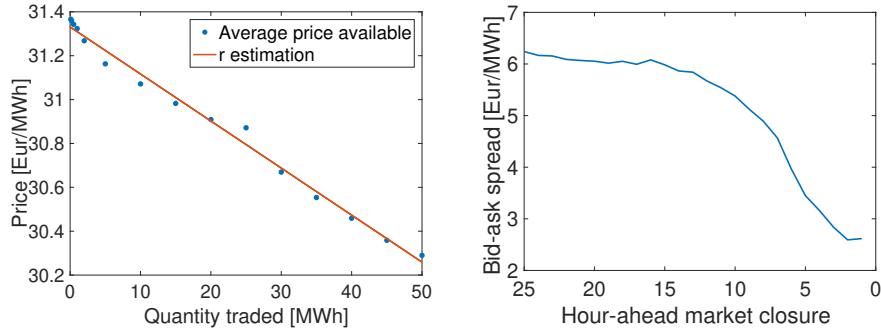


Figure 3.3: Estimation of r (left), and expected bid-ask-spread (right).

3.2 Analytical Solutions

3.2.1 Assumptions

In the analytical derivations of this section, we consider trading a fixed quantity of power. We assume the following:

- We discretize time. This approach is similar to the one proposed in [BP19b] and [SEM15].
- From one time step to the next one, the order book evolves as described in section 3.1.
- In order to simplify the analysis, we assume that the closer we are to market closure, the smaller the bid-ask spread is. This is consistent with the results that we obtain from the right panel of Fig. 3.3.
- We cannot be in imbalance at the end of the CIM. This assumption originates from the German regulation which discourages resources from being in imbalance on purpose [TEN].
- By considering every delivery period independently, we ignore time coupling effects. These have been considered in chapter 2.

3.2.2 Modelling the Problem as an MDP

As we trade a fixed quantity of power for a given delivery period, we can consider each delivery period independently. Therefore, our horizon ranges from the opening of the CIM to the closure of the CIM. We model our problem using the Markov Decision Process framework. To this aim, we need to define the state space, the action space, the reward function, as well as the state transition function.

Our state space contains two variables: (i) s_t , the quantity that still needs to be traded at time step t (for a trader in the CIM, it would correspond to the difference between the quantity that it had to trade initially and the quantity that it has already traded), and (ii) p_t , the center of the bid-ask-spread at time step t . Our action space consists of q_t , the quantity that we sell at time step t . The reward at time step t corresponds to the earnings obtained from selling q_t in the order book $p_t^s(q_t)$. It is expressed as:

$$\begin{aligned} rev(q_t) &= \int_0^{q_t} p_t^s(z) dz \\ &= \int_0^{q_t} (p_t - \Delta_t - 2rz) dz \\ &= p_t q_t - \Delta_t q_t - r q_t^2 \end{aligned}$$

The transition function links the state variable at time step $t - 1$ with the one at time step t :

$$s_{t-1} = s_t - q_t \quad (3.3)$$

$$p_{t-1} = p_t + \epsilon_t \quad (3.4)$$

Eq. (3.3) states that the quantity that needs to be traded at time step $t - 1$ is equal to the quantity that needs to be traded at time step t minus the quantity traded at time step t . Eq. (3.4) corresponds to the price evolution model of Eq. (3.2).

3.2.3 Optimal Trading Policy

In this section, we derive the optimal decision at each time step, as well as the optimal value function for the case in which we have a positive³ quantity to sell $s_t \geq 0$. We prove by induction, starting at the last time step 0, that the value function and the optimal decision at time step t are characterized by the formula in Table 3.1, where:

- $V_j^*(s_t, p_t) \doteq p_t s_t - \left(\sum_{i=0}^j \Delta_i \right) \frac{s_t}{j+1} - \frac{r s_t^2}{(j+1)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2$
- $C_j \doteq \frac{\sum_{i=0}^{j-1} \Delta_i - j \Delta_j}{2(j+1)r}$
- $X_j \doteq \left[\frac{j \Delta_j - \sum_{i=0}^{j-1} \Delta_i}{2r}, \frac{(j+1) \Delta_{j+1} - \sum_{i=0}^j \Delta_i}{2r} \right]$.

Range of quantity s_t to be traded	X_0	\dots	X_{t-1}	$R^+ \setminus (X_0 \cup \dots \cup X_{t-1})$
Optimal quantity to trade q_t^*	0	\dots	0	$\frac{s_t}{t+1} + \frac{\sum_{i=0}^{t-1} \Delta_i - t \Delta_t}{2(t+1)r}$
Value function $V_t^*(s_t, p_t)$	$V_0^*(s_t, p_t)$	\dots	$V_{t-1}^*(s_t, p_t)$	$p_t s_t - \left(\sum_{i=0}^t \Delta_i \right) \frac{s_t}{t+1} - \frac{r s_t^2}{(t+1)} + \sum_{i=1}^t \frac{(i+1)r}{i} C_i^2$

Table 3.1: Summary of the value function and optimal decision for step t .

We commence from time step 0. As explained in section 3.2.1, at the last time step we have to cover our position. Therefore, the decision is $q_0 = s_0$, and the associated value function is given by:

$$V_0^*(s_0, p_0) = p_0 s_0 - \Delta_0 s_0 - r s_0^2 \quad (3.5)$$

³The case with $s_t < 0$ can be computed similarly.

We now move to the first step of the induction. We first derive the optimal solution if we are one time step before delivery. The value function is given by:

$$\begin{aligned} V_1^*(s_1, p_1) &= \max_{0 \leq q_1 \leq s_1} p_1 q_1 - \Delta_1 q_1 - r(q_1)^2 + \int_{-\infty}^{\infty} V_0^*(s_0, p_0) f(p_0) dp_0 \\ &= \max_{0 \leq q_1 \leq s_1} p_1 q_1 - \Delta_1 q_1 - r(q_1)^2 + \int_{-\infty}^{\infty} (p_0 s_0 - \Delta_0 s_0 - r s_0^2) f(p_0) dp_0 \end{aligned} \quad (3.6)$$

$$\begin{aligned} &= \max_{0 \leq q_1 \leq s_1} p_1 q_1 - \Delta_1 q_1 - r(q_1)^2 + \mathbb{E}[p_0] s_0 - \Delta_0 s_0 - r s_0^2 \\ &= \max_{0 \leq q_1 \leq s_1} p_1 q_1 - \Delta_1 q_1 - r(q_1)^2 + p_1 s_0 - \Delta_0 s_0 - r s_0^2 \end{aligned} \quad (3.7)$$

$$\begin{aligned} &= \max_{0 \leq q_1 \leq s_1} p_1 q_1 - \Delta_1 q_1 - r(q_1)^2 + p_1(s_1 - q_1) - \Delta_0(s_1 - q_1) - r(q_1)^2 + 2r s_1 q_1 - r s_1^2 \\ &= \max_{0 \leq q_1 \leq s_1} \Delta_0(q_1 - s_1) - \Delta_1 q_1 - 2r(q_1)^2 + p_1 s_1 + 2r s_1 q_1 - r s_1^2 \end{aligned} \quad (3.8)$$

For Eq. (3.6), we use the definition of the value function (3.5). In Eq. (3.7), we use the fact that $\mathbb{E}[p_0] = p_1$. For Eq. (3.8), we use the transition function of s_t defined in (3.3). As this objective function is quadratic, we can compute the maximum if we would be ignoring the constraints. To this aim, we compute the point at which the gradient vanishes:

$$\begin{aligned} 0 &= \Delta_0 - \Delta_1 - 4r q_1^* + 2r s_1 \\ 4r q_1^* &= \Delta_0 - \Delta_1 + 2r s_1 \\ q_1^* &= \frac{s_1}{2} + \frac{\Delta_0 - \Delta_1}{4r} \end{aligned}$$

There are three cases to consider for this maximum:

- **The maximum is feasible, $0 \leq q_1^* \leq s_1$:** This condition is equivalent to $\frac{\Delta_1 - \Delta_0}{2r} \leq s_1$. In this case, we have the optimal decision. The associated value function is given by the following expression, in which we fix $C_1 = \frac{\Delta_0 - \Delta_1}{4r}$:

$$\begin{aligned} V_1^*(s_1, p_1) &= \Delta_0 \left(-\frac{s_1}{2} + C_1 \right) - \Delta_1 \left(\frac{s_1}{2} + C_1 \right) - 2r \left(\frac{s_1}{2} + C_1 \right)^2 \\ &\quad + p_1 s_1 + 2r s_1 \left(\frac{s_1}{2} + C_1 \right) - r s_1^2 \\ &= \Delta_0 \left(-\frac{s_1}{2} + C_1 \right) - \Delta_1 \left(\frac{s_1}{2} + C_1 \right) - \frac{r}{2} s_1^2 - 2r C_1^2 + p_1 s_1 \\ &= -\frac{\Delta_0}{2} s_1 - \frac{\Delta_1}{2} s_1 - \frac{r}{2} s_1^2 + p_1 s_1 + C_1 (\Delta_0 - \Delta_1 - 2r C_1) \\ &= -\frac{\Delta_0}{2} s_1 - \frac{\Delta_1}{2} s_1 - \frac{r}{2} s_1^2 + p_1 s_1 + 2r C_1^2 \end{aligned}$$

- $q_1^* < 0$: This condition is equivalent to $s_1 \leq \frac{\Delta_1 - \Delta_0}{2r}$. In this case, we observe that the derivative is always negative, because (i) $\Delta_0 - \Delta_1 + 2r s_1$ is negative, and (ii) $-4r q_1$ is negative for $q_1 \geq 0$. Therefore, it is optimal to have $q_1^* = 0$. This case represents the situation in which the spread at the last time step is significantly smaller than the one at time step 1 ($\Delta_1 \gg \Delta_0$). Therefore, it is optimal to trade the entire remaining quantity at the last time step. The associated value function is given by:

$$V_1^*(s_1, p_1) = p_1 s_1 - \Delta_0 s_1 - r s_1^2$$

- $q_1 > s_1$: This solution can never occur because, as assumed in section 3.2.1, $\Delta_0 \leq \Delta_1$.

We summarize the value function of time step 1 as a function of the quantity to trade s_1 and the price p_1 in Table 3.2.

Range of quantity s_1 to be traded	$[0, \frac{\Delta_1 - \Delta_0}{2r} [$	$[\frac{\Delta_1 - \Delta_0}{2r}, \infty [$
Optimal quantity to trade q_1^*	0	$\frac{s_1}{2} + \frac{\Delta_0 - \Delta_1}{4r}$
Value function $V_1^*(s_1, p_1)$	$(p_1 - \Delta_0 - r s_1) s_1$	$(p_1 - \frac{\Delta_0}{2} - \frac{\Delta_1}{2} - \frac{r}{2} s_1) s_1 + 2r C_1^2$

Table 3.2: Summary of the value function and optimal decision for step 1.

In order to conclude the induction proof, we need to prove that, if the value function and the optimal quantity to trade at time step t follows the format given in Table 3.3, the optimal quantity and the associated value function have the same format at time step $t + 1$ as shown in Table 3.4 where:

- $V_j(s_t, p_t) = p_t s_t - \left(\sum_{i=0}^j \Delta_i \right) \frac{s_t}{j+1} - \frac{r s_t^2}{(j+1)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2$
- $C_j = \frac{\sum_{i=0}^{j-1} \Delta_i - j \Delta_j}{2(j+1)r}$
- $X_j = \left[\frac{j \Delta_j - \sum_{i=0}^{j-1} \Delta_i}{2r}, \frac{(j+1) \Delta_{j+1} - \sum_{i=0}^j \Delta_i}{2r} \right[$.

Range of quantity s_t to be traded	X_0	\dots	X_{t-1}	$R^+ \setminus (X_0 \cup \dots \cup X_{t-1})$
Optimal quantity to trade q_t^*	0	\dots	0	$\frac{s_t}{t+1} + \frac{\sum_{i=0}^{t-1} \Delta_i - t \Delta_t}{2(t+1)r}$
Value function $V_t^*(s_t, p_t)$	$V_0^*(s_t, p_t)$	\dots	$V_{t-1}^*(s_t, p_t)$	$p_t s_t - \left(\sum_{i=0}^t \Delta_i \right) \frac{s_t}{t+1} - \frac{r s_t^2}{(t+1)} + \sum_{i=1}^t \frac{(i+1)r}{i} C_i^2$

Table 3.3: Summary of the value function and optimal decision for step t .

Range of quantity s_{t+1} to be traded	X_0	\dots	X_t	$R^+ \setminus (X_0 \cup \dots \cup X_t)$
Optimal quantity to trade q_{t+1}^*	0	\dots	0	$\frac{s_{t+1}}{t+2} + \frac{\sum_{i=0}^t \Delta_i - (t+1) \Delta_{t+1}}{2(t+2)r}$
Value function $V_{t+1}(s_{t+1}, p_{t+1})$	$V_0^*(s_{t+1}, p_{t+1})$	\dots	$V_t^*(s_{t+1}, p_{t+1})$	$p_{t+1} s_{t+1} - \left(\sum_{i=0}^{t+1} \Delta_i \right) \frac{s_{t+1}}{t+2} - \frac{r s_{t+1}^2}{(t+2)} + \sum_{i=1}^{t+1} \frac{(i+1)r}{i} C_i^2$

Table 3.4: Summary of the value function and optimal decision for step $t + 1$

We start the proof by observing that, as shown in Table 3.3, the value function at time step t can be decomposed in $t + 1$ different options depending of the quantity that still needs to be traded, s_t . Therefore, we can write the value function at time step $t + 1$ as the maximum of $t + 1$ functions:

$$V_{t+1}^*(s_{t+1}, p_{t+1}) = \max \left(V_{t+1,0}^*(s_{t+1}, p_{t+1}), \dots, V_{t+1,t}^*(s_{t+1}, p_{t+1}) \right)$$

where $V_{t+1,j}^*(s_{t+1}, p_{t+1})$ corresponds to the optimal value function at time step $t + 1$ under the constraint that $s_{t+1} - q_{t+1} \in X_j$ (option j is optimal at time step t). Mathematically, this can be written as:

$$V_{t+1,j}^*(s_{t+1}, p_{t+1}) = \max_{q_{t+1} \in Y_j} p_{t+1} q_{t+1} - \Delta_{t+1} q_{t+1} - r(q_{t+1})^2 + V_j^*(s_{t+1} - q_{t+1}, p_{t+1}) \quad (3.9)$$

We refer to this as the value function for option j , at time step $t + 1$. The interval Y_j is defined as the interval of the quantity q_{t+1} such that, if we trade q_{t+1} , the remaining quantity to trade at time step t , s_t , is in the interval X_j . Mathematically, this can be expressed as:

- Y_0 is the interval $[\max(s_{t+1} - \frac{\Delta_1 - \Delta_0}{2r}, 0), s_{t+1}]$.
- $Y_j, \forall j \in 1..t-1$, is the interval

$$[\max(0, s_{t+1} - \frac{(j+1)\Delta_{j+1} - \sum_{i=0}^j \Delta_i}{2r}), \min(s_{t+1} - \frac{j\Delta_j - \sum_{i=0}^{j-1} \Delta_i}{2r}, s_{t+1})]$$

As the spread is decreasing⁴, $s_{t+1} \geq s_{t+1} - \frac{j\Delta_j - \sum_{i=0}^{j-1} \Delta_i}{2r}$. Therefore, Y_j can be simplified as:

$$[\max(0, s_{t+1} - \frac{(j+1)\Delta_{j+1} - \sum_{i=0}^j \Delta_i}{2r}), s_{t+1} - \frac{j\Delta_j - \sum_{i=0}^{j-1} \Delta_i}{2r}] \quad (3.10)$$

- Y_t is the interval

$$[0, \min(s_{t+1} - \frac{t\Delta_t - \sum_{i=0}^{t-1} \Delta_i}{2r}, s_{t+1})]$$

As the spread is decreasing, $s_{t+1} \geq s_{t+1} - \frac{t\Delta_t - \sum_{i=0}^{t-1} \Delta_i}{2r}$. Therefore, Y_t can be simplified as:

$$[0, s_{t+1} - \frac{t\Delta_t - \sum_{i=0}^{t-1} \Delta_i}{2r}]$$

The idea of the rest of the proof is to compare the value function of the different options $V_{t+1,j}$ depending of the quantity to trade s_{t+1} . We start by computing the value of s_{t+1} for which the interval Y_j is empty:

- $s_{t+1} \in X_0$ ($0 \leq s_{t+1} \leq \frac{\Delta_1 - \Delta_0}{2r}$):

In this case, $Y_j, \forall j \in 1 \dots t$, is empty because its upper bound (Eq.(3.10)) is negative. Therefore, there is only one possible option: $V_{t+1,0}$.

- $s_{t+1} \in R \setminus X_0 \cup \dots \cup X_{t-1}$ ($s_{t+1} \geq \frac{t\Delta_t - \sum_{i=0}^{t-1} \Delta_i}{2r}$):

In this case, the upper bound of Y_0, \dots, Y_t are positive. Therefore, all the options are feasible.

- $s_{t+1} \in X_k$ ($\frac{k\Delta_k - \sum_{i=0}^{k-1} \Delta_i}{2r} \leq s_{t+1} \leq \frac{(k+1)\Delta_{k+1} - \sum_{i=0}^k \Delta_i}{2r}$):

In this case, $V_{t+1,j}, \forall j \in k+1 \dots t$ is infeasible because the upper bound (Eq.(3.10)) of Y_j is negative in interval X_k . Therefore, the only feasible options are $V_{t+1,j}, \forall j \in 0 \dots k$.

Table 3.5 summarizes the value functions that are infeasible for the different options in the different intervals of s_{t+1} .

In order to compute the value functions associated to the feasible options, we use 3 propositions, which we derive below. These propositions compute the value function evaluated at the optimal unconstrained quantity and at a quantity equal to 0. The proof of these propositions can be found in appendix 3.6.

Proposition 1. *The unconstrained optimal solution of $V_{t+1,j}^*(s_{t+1}, p_{t+1})$ is given by:*

$$q_{t+1,j}^* = \frac{s_{t+1}}{j+2} - \frac{(j+1)\Delta_{t+1} - \left(\sum_{i=0}^j \Delta_i\right)}{2r(j+2)}$$

⁴The closer we are to market closure, the smaller the bid-ask spread is.

s_{t+1}	X_0	X_1	\cdots	X_{j-1}	X_j	X_{j+1}	\cdots	X_t	$R^+ \setminus (X_0 \cup \cdots \cup X_t)$
$V_{t+1,0}$									
$V_{t+1,1}$	$-\infty$								
\vdots	\ddots	\ddots							
$V_{t+1,j-1}$	$-\infty$	$-\infty$	\ddots						
$V_{t+1,j}$	$-\infty$	$-\infty$	\ddots	$-\infty$					
$V_{t+1,j+1}$	$-\infty$	$-\infty$	\ddots	$-\infty$	$-\infty$				
\ddots	\ddots	\ddots	\ddots	\ddots	\ddots	\ddots			
$V_{t+1,t}$	$-\infty$	$-\infty$	\ddots	$-\infty$	$-\infty$	$-\infty$	\ddots		

Table 3.5: Value function for the different options in the different intervals of s_{t+1} .

Proposition 2. *The value function evaluated at this optimal unconstrained solution $q_{t+1,j}^*$, defined as $V_{t+1,j}^u(s_{t+1}, p_{t+1})$, is given by:*

$$\begin{aligned}
V_{t+1,j}^u(s_{t+1}, p_{t+1}) &= p_{t+1}s_{t+1} - \Delta_{t+1} \frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} \\
&\quad + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 + r \frac{j+2}{j+1} C_{t+1,j}^2
\end{aligned} \tag{3.11}$$

where $C_{t+1,j} = \frac{(j+1)\Delta_{t+1} - (\sum_{i=0}^j \Delta_i)}{2r(j+2)}$

Proposition 3. *The value function evaluated at $q_{t+1,j} = 0$ defined as $V_{t+1,j}^0(s_{t+1}, p_{t+1})$ is given by:*

$$\begin{aligned}
V_{t+1,j}^0(s_{t+1}, p_{t+1}) &= p_{t+1}s_{t+1} - \frac{s_{t+1}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \left(\frac{s_{t+1}^2}{j+1} \right) \\
&\quad + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2
\end{aligned} \tag{3.12}$$

Using these 3 propositions, we compute the value function for the different options j depending on the quantity to trade s_{t+1} :

- $s_{t+1} \in X_0$ ($0 \leq s_{t+1} \leq \frac{\Delta_1 - \Delta_0}{2r}$)

For option 0, the optimal unconstrained quantity is given by:

$$q_{t+1,0}^* = \frac{s_{t+1}}{2} - \frac{\Delta_1 - \Delta_0}{4r}$$

as proven in proposition 1. By the definition of X_0 , the optimal unconstrained quantity is negative. Therefore, the optimal quantity to trade is equal to 0 and the associated value function is given by V_{t+1}^0 (proposition 3).

- $s_{t+1} \in R \setminus (X_0 \cup \cdots \cup X_t)$ ($s_{t+1} \geq \frac{(t+1)\Delta_{t+1} - \sum_{i=0}^t \Delta_i}{2r}$)

For option t , the optimal unconstrained quantity is given by:

$$q_{t+1,t}^* = \frac{s_{t+1}}{t+2} - \frac{(t+1)\Delta_{t+1} - \sum_{i=0}^t \Delta_i}{2r(t+2)}$$

as proven in proposition 1. By definition of the interval, this optimal unconstrained quantity is feasible. For option j , $\forall j \in 0 \cdots t-1$, we use the value function evaluated at the unconstrained optimal quantity $V_{t+1,j}^u$ as an upper bound for the optimal value function $V_{t+1,j}^*$.

- $s_{t+1} \in X_k$ ($\frac{k\Delta_k - \sum_{i=0}^{k-1} \Delta_i}{2r} \leq s_{t+1} \leq \frac{(k+1)\Delta_{k+1} - \sum_{i=0}^k \Delta_i}{2r}$)

For option k , the optimal unconstrained quantity is given by:

$$q_{t+1,k}^* = \frac{s_{t+1}}{k+2} - \frac{(k+1)\Delta_{t+1} - \left(\sum_{i=0}^k \Delta_i\right)}{2r(k+2)}$$

as proven in proposition 1. By definition of X_k , the optimal unconstrained quantity is negative. Therefore, the optimal quantity to trade is equal to 0. For option j , $\forall j \in 0 \cdots k-1$, we define an upper bound to the optimal value function $V_{t+1,j}^*$ as:

$$\bar{V}_{t+1,j} = \begin{cases} V_{t+1,j}^u & \text{If } q_{t+1,j}^* > 0 \\ V_{t+1,j}^0 & \text{If } q_{t+1,j}^* \leq 0 \end{cases}$$

Table 3.6 presents the value function of the different options in the different intervals of s_{t+1} . The value functions that have been computed exactly are highlighted in blue. The other terms are the upper bounds for the other options.

s_{t+1}	X_0	X_1	\cdots	X_{j-1}	X_j	X_{j+1}	\cdots	X_t	$R^+ \setminus (X_0 \cup \cdots \cup X_t)$
$V_{t+1,0}$	$V_{t+1,0}^0$	$\bar{V}_{t+1,0}$	\cdots	$\bar{V}_{t+1,0}$	$\bar{V}_{t+1,0}^u$	$\bar{V}_{t+1,0}$	\cdots	$\bar{V}_{t+1,0}$	$V_{t+1,0}^u$
$V_{t+1,1}$	$-\infty$	$V_{t+1,1}^0$	\ddots	$\bar{V}_{t+1,1}$	$\bar{V}_{t+1,1}$	$\bar{V}_{t+1,1}$	\ddots	$\bar{V}_{t+1,1}$	$V_{t+1,1}^u$
\vdots	\ddots	\ddots	\ddots	\ddots	\ddots	\ddots	\ddots	\ddots	\vdots
$V_{t+1,j-1}$	$-\infty$	$-\infty$	\ddots	$V_{t+1,j-1}^0$	$\bar{V}_{t+1,j-1}$	$\bar{V}_{t+1,j-1}$	\ddots	$\bar{V}_{t+1,j-1}$	$V_{t+1,j-1}^u$
$V_{t+1,j}$	$-\infty$	$-\infty$	\ddots	$-\infty$	$V_{t+1,j}^0$	$\bar{V}_{t+1,j}$	\ddots	$\bar{V}_{t+1,j}$	$V_{t+1,j}^u$
$V_{t+1,j+1}$	$-\infty$	$-\infty$	\ddots	$-\infty$	$-\infty$	$V_{t+1,j+1}^0$	\vdots	$\bar{V}_{t+1,j+1}$	$V_{t+1,j+1}^u$
\vdots	\ddots	\ddots	\ddots	\ddots	\ddots	\ddots	\ddots	\ddots	\vdots
$V_{t+1,t}$	$-\infty$	$-\infty$	\ddots	$-\infty$	$-\infty$	$-\infty$	\ddots	$V_{t+1,t}^0$	$V_{t+1,t}^u$

Table 3.6: Value function for the different options in the different intervals of s_{t+1}

In order to conclude the proof, we show that the exact value functions are higher than the upper bound for the other options. We use the following three propositions⁵ to compare the value function of consecutive options:

Proposition 4. For two consecutive unconstrained value functions $V_{t+1,j-1}^u(s_{t+1}, p_{t+1})$ and $V_{t+1,j}^u(s_{t+1}, p_{t+1})$,

$$V_{t+1,j}^u(s_{t+1}, p_{t+1}) \geq V_{t+1,j-1}^u(s_{t+1}, p_{t+1}), \quad \forall s_{t+1}, \forall p_{t+1}$$

Proposition 5. For the value function of option j evaluated at 0, $V_{t+1,j}^0(s_{t+1}, p_{t+1})$, and the value function of option $j-1$ evaluated at the unconstrained optimal quantity $V_{t+1,j-1}^u(s_{t+1}, p_{t+1})$,

$$V_{t+1,j}^0(s_{t+1}, p_{t+1}) \geq V_{t+1,j-1}^u(s_{t+1}, p_{t+1}), \quad \forall p_{t+1}, \forall s_{t+1} \geq \frac{j(\Delta_j + \Delta_{t+1}) - 2\left(\sum_{i=0}^{j-1} \Delta_i\right)}{4r}$$

⁵These propositions have been proven in appendix 3.7.

Proposition 6. For two consecutive value functions evaluated at 0, $V_{t+1,j-1}^0(s_{t+1}, p_{t+1})$ and $V_{t+1,j}^0(s_{t+1}, p_{t+1})$,

$$V_{t+1,j}^0(s_{t+1}, p_{t+1}) \geq V_{t+1,j-1}^0(s_{t+1}, p_{t+1}), \forall p_{t+1}, \forall s_{t+1}$$

We compare the value function for the different options depending on the quantity to trade s_{t+1} :

- $s_{t+1} \in R^+ \setminus (X_0 \cup \dots \cup X_t)$:

The unconstrained optimal quantity is feasible for option t . Therefore, we can apply proposition 4 recursively in order to obtain:

$$V_{t+1,t}^u(s_{t+1}, p_{t+1}) \geq V_{t+1,t-1}^u(s_{t+1}, p_{t+1}) \geq \dots \geq V_{t+1,1}^u(s_{t+1}, p_{t+1}) \geq V_{t+1,0}^u(s_{t+1}, p_{t+1})$$

We conclude that the option with the highest value function is t and that the associated value function is:

$$\begin{aligned} V_{t+1,t}^*(s_{t+1}, p_{t+1}) &= p_{t+1}s_{t+1} - \Delta_{t+1} \frac{s_{t+1}}{t+2} - \frac{s_{t+1}}{t+2} \left(\sum_{i=0}^t \Delta_i \right) - r \frac{s_{t+1}^2}{(t+2)} \\ &\quad + \sum_{i=1}^t \frac{(i+1)r}{i} C_i^2 + r \frac{t+2}{t+1} C_{t+1,t}^2 \\ &= p_{t+1}s_{t+1} - \Delta_{t+1} \frac{s_{t+1}}{t+2} - \frac{s_{t+1}}{t+2} \left(\sum_{i=0}^t \Delta_i \right) - r \frac{s_{t+1}^2}{(t+2)} \\ &\quad + \sum_{i=1}^{t+1} \frac{(i+1)r}{i} C_i^2 \end{aligned}$$

- $s_{t+1} \in X_j, \forall j \in 1 \dots t-1$:

We start by comparing the value function associated to option j , $V_{t+1,j}^0$ and the upper bound on the value function associated to option $j-1$, $\bar{V}_{t+1,j-1}$. To this aim, we compute the unconstrained optimal quantity associated to option $j-1$ using proposition 1:

$$q_{t+1,j-1}^* = \frac{s_{t+1}}{j+1} - \frac{j\Delta_{t+1} - \left(\sum_{i=0}^{j-1} \Delta_i \right)}{2r(j+1)}$$

From this unconstrained optimal quantity, we can distinguish two cases:

1. **When the unconstrained optimal quantity is positive** ($q_{t+1,j-1}^* \geq 0$):

In this case, we have that

$$s_{t+1} \geq \frac{j\Delta_{t+1} - \left(\sum_{i=0}^{j-1} \Delta_i \right)}{2r} \quad (3.13)$$

Using proposition 5, we know that:

$$V_{t+1,j}^0(s_{t+1}, p_{t+1}) \geq V_{t+1,j-1}^u(s_{t+1}, p_{t+1}), \forall p_{t+1}, \forall s_{t+1} \geq \frac{j(\Delta_j + \Delta_{t+1}) - 2 \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r} \quad (3.14)$$

It can be observed that, if $q_{t+1,j-1}^*$ is positive, the condition of proposition 5 is also respected because the lower bound of this condition (Eq. (3.14)) is lower⁶ than the bound for $q_{t+1,j-1}^*$ to be positive (Eq. (3.13)):

$$\begin{aligned} \frac{j(\Delta_j + \Delta_{t+1}) - 2 \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r} &\leq \frac{2j\Delta_{t+1} - 2 \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r} \\ &= \frac{j\Delta_{t+1} - \left(\sum_{i=0}^{j-1} \Delta_i \right)}{2r} \end{aligned}$$

This establishes that the value function associated to option j evaluated at 0 is higher than the value function associated to option $j - 1$ evaluated at the optimal unconstrained solution.

2. **When the unconstrained optimal quantity is negative** ($q_{t+1,j-1}^* < 0$):
In this case, we know from proposition 6 that:

$$V_{t+1,j}^0(s_{t+1}, p_{t+1}) \geq V_{t+1,j-1}^0(s_{t+1}, p_{t+1})$$

This establishes that the value function associated to option j evaluated at 0 is higher than the value function associated to option $j - 1$ evaluated at 0.

The same reasoning can be applied in order to prove that the value function associated to option $j - 1$ is higher than the one associated to option $j - 2$. By applying this recursively, we obtain that the value function associated to option j is the highest one in interval X_j . The optimal decision is to sell 0 and the optimal value function is given by:

$$\begin{aligned} V_{t+1}^*(s_{t+1}, p_{t+1}) &= V_{t+1,j}^0(s_{t+1}, p_{t+1}) \\ &= p_{t+1}s_{t+1} - \frac{s_{t+1}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \left(\frac{s_{t+1}^2}{j+1} \right) + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \end{aligned}$$

The final results are summarized in Table 3.7. This table is exactly the same as Table 3.4, which concludes the proof.

3.2.4 Insights from the Analytical Solution

In this section, we present the insights that we gain from the analytical solution.

⁶This is due to the fact that Δ_{t+1} is bigger than Δ_j . The reason for this is that we assume that the closer we are from market closure, the smaller the bid-ask spread is.

Range of quantity s_{t+1} to be traded	X_0	\dots	X_t	$R^+ \setminus (X_0 \cup \dots \cup X_t)$
Optimal quantity to trade q_{t+1}^*	0	\dots	0	$\frac{s_{t+1}}{t+2} + \frac{\sum_{i=0}^t \Delta_i - (t+1)\Delta_{t+1}}{2(t+2)r}$
Value function $V_{t+1}(s_{t+1}, p_{t+1})$	$V_0^*(s_{t+1}, p_{t+1})$	\dots	$V_t^*(s_{t+1}, p_{t+1})$	$p_{t+1}s_{t+1} - \left(\sum_{i=0}^{t+1} \Delta_i \right) \frac{s_{t+1}}{t+2} - \frac{rs_{t+1}^2}{(t+2)} + \sum_{i=1}^{t+1} \frac{(i+1)r}{i} C_i^2$

Table 3.7: Summary of the value function and optimal decision for step $t + 1$

Optimal quantity to trade The optimal quantity to trade at time step t (there are still $t + 1$ chances to trade) is given by:

$$\frac{s_t}{t+1} + \frac{\sum_{i=0}^{t-1} \Delta_i - t\Delta_t}{2(t+1)r}.$$

If we would consider only the first term, it would imply that we trade $\frac{1}{t+1}$ of the capacity available. The second term is a correction for the difference in spread between the different time steps. This term is negative, which means that we always trade at most $\frac{s_t}{t+1}$. We can analyse two extreme cases (i) $r \rightarrow 0$: This means that the second term is very negative. In this situation, the optimal decision is to always trade 0 until the last time step. This results from the fact that the price impact is negligible, and therefore we can trade all the power when the spread is the lowest. (ii) *The spread is constant*: This means that the second term is equal to 0. In this case, it is optimal to trade the same quantity at each time step. This can be interpreted as follows: As the spread is constant, there is no reason to prefer one time step over another and therefore we simply aim at minimizing our impact on the price. Another property of the optimal decision that is worth highlighting is that it does not depend on the middle of the bid-ask-spread, p_t . This property stems from the assumption that we adopt on the evolution of the center of the bid-ask spread in (Eq. 3.4). Indeed, no matter the value of p_t , we assume that $\mathbb{E}[p_{t-1}] = p_t$. Therefore, there is no reason for us to sell more (resp. less) power at the current time step, in anticipation that the the center of the bid-ask-spread will become lower (resp. higher) than the value that we currently observe.

Intuition about the value function The optimal value function is presented in Table 3.3. From this table, we observe that for different ranges of power to sell, s_t , we have different expressions for the value function. These different ranges represent different time steps over which we are required to trade. For instance, (i) X_0 represents the case for which we have a small quantity of power to sell. In this case, we can ignore the price impact and trade all the power at the last time step where the bid-ask-spread is the lowest. (ii) X_1 represents the case for which we have more power to sell. In this situation, we would be affected more significantly by the price impact, and we therefore split the quantity that we have to sell between the two last time steps.

Mathematical format of the value function As explained before, the value function is a piecewise function. For each of these pieces, the value function can be decomposed as (i) a quadratic function in s_t , and (ii) a bilinear term in $p_t s_t$. We exploit this observation in the algorithmic section.

Example with the true estimated parameters We illustrate, in Fig. 3.4, the trading behaviour that we obtain if we apply the analytical solution to the simplified market with the parameters that are estimated in section 3.1. More precisely, we present the quantity

that still needs to be traded s_t for each time step. In the left panel, we present the results if we start with 25MWh to sell 10 hours before market closure. In this situation, we observe that we only start trading 2 hours before delivery. This means that, for that range of power, the difference in bid-ask-spread is more important to consider than the price impact. In the right graph, we present the results if we start with 100MWh to sell 10 hours before market closure. As expected by the analytical solution, we start trading earlier (4 hours ahead) than if we have 25MWh to trade. This illustrates that, when we have a larger quantity of power to sell, it is not sufficient to trade all the power at the moment at which the bid-ask-spread is the smallest, we also need to account for our impact on the price while we are trading.

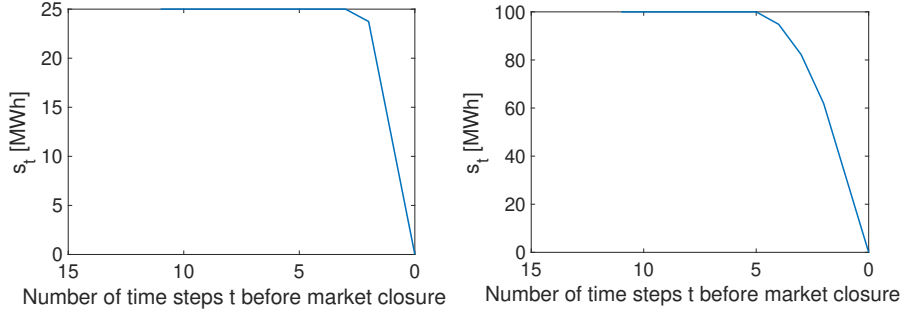


Figure 3.4: Analytical solution: quantity that remains to be traded, s_t , with an initial quantity of 25MWh (left) and 100MWh (right).

3.3 Algorithmic Approaches

We now use the modeling setup and the insights of the analytical solution as a basis for a VFA and SDDP algorithm.

3.3.1 Value Function Parametrization

In our problem, we face continuous state and action spaces. In order to develop a VFA algorithm, we parametrize the action-value function $q^*(s, a)$ as $\hat{q}(s, a; w)$, where w is a set of parameters that we need to optimize.

In order to determine basis functions for the value function, we exploit the insights from the analytical solution. By definition of the action-value function [SB18], we know that:

$$Q_t((s_t, p_t), q_t) = p_t q_t - \Delta_t q_t - r(q_t)^2 + V_{t-1}^*(s_t - q_t, p_t)$$

where $V_{t-1}^*(s_t - q_t, p_t)$ is the optimal value function described in section 3.2.3. The parametrization of this action-value function can therefore be split into two parts:

$$\hat{q}((s_t, p_t), q_t; w) = f_0(s_t, p_t, q_t; w_0) + \sum_{j=1}^T f_j(s_t - q_t, p_t; w_j) \cdot \mathbf{1}_j$$

In this expression:

- f_0 contains the parameters for the payoff obtained at time step t which can be parametrized as:

$$f_0(s_t, p_t, q_t; w_0) = w_{0,0} \cdot (p_t - \Delta_t) \cdot q_t + w_{0,1} \cdot q_t^2.$$

- f_j is the basis function that represents $V_j^*(s_t - q_t, p_t)$. As we know that this function is piecewise quadratic in $s_t - q_t$ and has a bilinear term in $p_t(s_t - q_t)$, we can parametrize it as:

$$f_j(s_t - q_t, p_t; w_j) = w_{j,0} + w_{j,1} \cdot (s_t - q_t) + w_{j,2} \cdot (s_t - q_t)^2 + w_{j,3} \cdot p_t(s_t - q_t).$$

- $\mathbf{1}_j$ is an indicator function which evaluates to 1 if f_j should be used and 0 otherwise. Based on Table 3.3, there are two cases in which f_j should be used (i) $t > j$ and $s_t - q_t \in X_j$, and (ii) $t = j$ and $s_t - q_t \in R^+ \setminus (X_0 \cup \dots \cup X_{j-1})$.

Having defined basis functions, we minimize the error between the prediction from our value function $\hat{q}((s_t, p_t), q_t; w)$ and the obtained outcome from the episode, using the update presented in section 1.4.3.

3.3.2 Stochastic Dual Dynamic Programming

SDDP is a method for solving a specific class of multi-stage stochastic convex programs. For a complete description of the algorithm, the reader can refer to [PP91]. In order to use SDDP, we need to define the subproblem faced at every stage. This subproblem, referred to as the Nested L-Shaped Decomposition Subproblem (NLDS), is defined for every time step t and node k of the uncertainty model. In our case, the NLDS is expressed as:

$$\begin{aligned} \max_{q_t^s \geq 0, q_t^b \geq 0} \quad & (p_{t,k} - \Delta_t - r q_t^s) q_t^s - (p_{t,k} + \Delta_t + r q_t^b) q_t^b \\ \text{s.t.} \quad & s_t = s_{t-1} - q_t^s + q_t^b \end{aligned}$$

There are two specificities compared to the classical problems solved using SDDP. (i) Our objective function is quadratic in q_t^s and q_t^b . These terms can be linearized using a first-order Taylor approximation [BLS⁺12]. (ii) The uncertain parameter $p_{t,k}$ appears in the objective function and follows an auto-regressive process, based on our model for the price evolution (Eq. (3.2)). This is not suitable for classical implementations of SDDP [DDB20] because the value function is convex in s_t and concave in p_t . Therefore, we use the modified version of the algorithm that is presented in [DDB20] and implemented in toolbox [DK20]. As explained in [DDB20], this algorithm is guaranteed to converge almost surely to the optimal solution.

3.4 Case Study

In this section, we compare the main results obtained by the analytical solution, VFA and SDDP. We start by comparing these methods based on their numerical performance. Afterwards, we present their run-times.

Numerical results: We compare the numerical results obtained by the analytical solution, VFA and SDDP for a problem with 10 time steps. We start by comparing the evolution of the value function with respect to the initial quantity to trade, in the left graph of Fig. 3.5. We observe that both SDDP and VFA provide very close results compared to the analytical solution. To be more specific, we present, in the right panel of Fig. 3.5, the error in the value function computed by ADP and SDDP relative to the

value function of the analytical solution. We observe that these differences are very small (around 1 euro) compared to the magnitude of the value function (around 1000 euros). Finally, we present, in Fig. 3.6, the evolution of the quantity s_t that still needs to be traded for each time time step. We observe that the decisions that are reached by SDDP and VFA are very close to the ones obtained by the analytical solution. This confirms that both methods are suitable for solving this simplified problem.

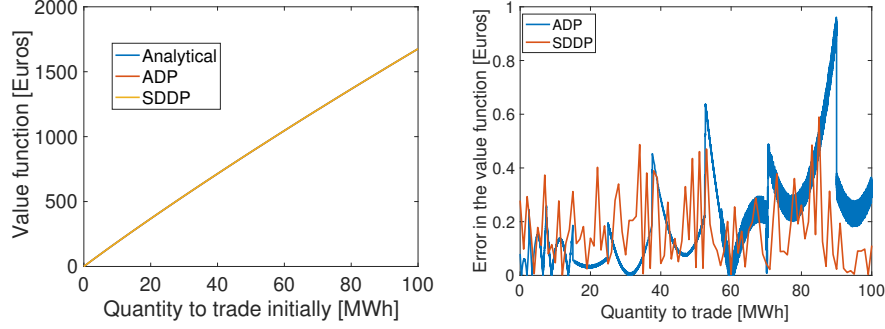


Figure 3.5: comparison of value functions for the analytical solution, VFA and SDDP (left). Error in the value function of ADP and SDDP compared to the analytical solution (right).

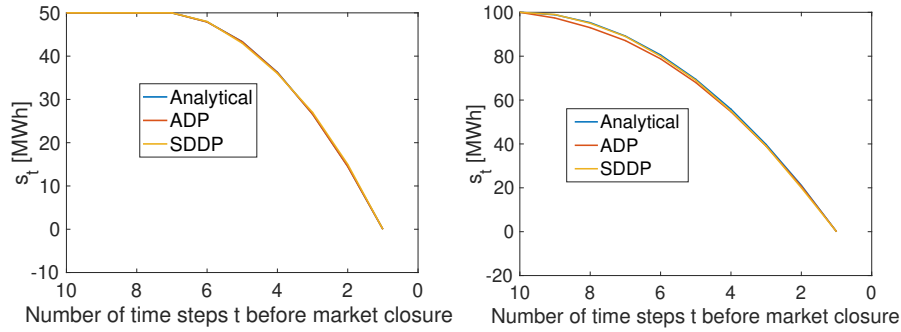


Figure 3.6: Quantity that still needs to be traded, s_t , with an initial quantity of 50MWh (left) and 100MWh (right).

Run-times Regarding the run-time required when applying the method in real time, it is negligible for all methods, since the work has been performed off-line. For the analytical solution and for VFA, we simply need to evaluate a number of functions (there are more function evaluations in VFA than for the analytical solution). For SDDP, we need to solve one NLDS, which is a linear program. This is slightly more demanding than evaluating functions, but still fairly fast. The second run-time to consider is the one for performing the offline computation. We compare SDDP to ADP with respect to this run time on the example with 10 time steps in Fig. 3.7. Specifically, we run the learning process for 21 different quantities to trade (uniformly distributed between 0 and 100MW). We compute the error in the value function after a certain time and plot the average error in Fig. 3.7. From this figure, we observe that SDDP seems faster than VFA. For instance, it reaches a precision of 1% in 4 seconds while VFA requires approximately 40 seconds for reaching the same precision.

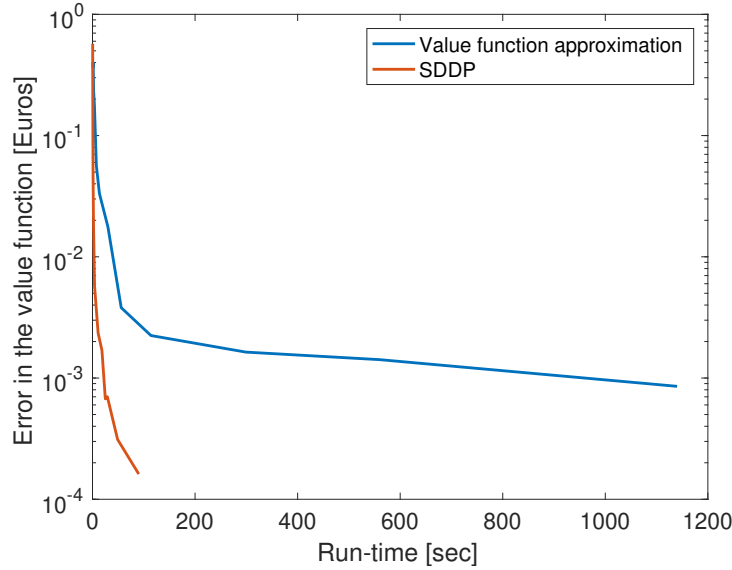


Figure 3.7: Comparison of the error percentage in the value function with respect to the run-time for VFA and SDDP.

3.5 Conclusions and Perspectives

In this chapter, we model the problem of selling a fixed quantity of power in a simplified model of the CIM using the MDP framework. We derive the optimal trading strategy for this problem through backward induction. We use the optimal value function to develop basis functions for a VFA algorithm. We then develop VFA and SDDP algorithms for testing against our idealized CIM model and demonstrate that they both arrive to the optimal policy on a 10-step example.

In future work, we aim at enriching this algorithmic framework in order to trade the output of a renewable unit in the real CIM. To this aim, we would need to adapt the algorithms in the following ways. Firstly, we need to be able to include uncertainty in the production output. The model would also need to be adapted to the fact that the CIM price evolution does not follow precisely the model of section 1.3.3. Deviations, for example, may result from the fact that, in practice, r and Δ_t do not evolve deterministically. Moreover, it is possible that the center of the bid-ask spread does not obey Eq. (3.2). Indeed, in practice, the information available in the CIM may allow us to predict if the CIM price would increase or decrease on average.

For the analytical solution, it is clear that these changes would require revisiting the analytical framework. For SDDP, the uncertainty on the unit production can be represented by modifying the NLDS. We can even consider that the output of our renewable unit is (negatively) correlated with the center of the bid-ask-spread. This is likely the case, because if we over-produce (resp. under-produce), there is a high chance that other producers will also over-produce (resp. under-produce), which would push the price down (resp. up). Another adaptation that can be considered with respect to SDDP is to assume that the evolution of the r and Δ_t parameters is stochastic. These parameters can even be considered as (positively) correlated. An illiquid (resp. liquid) market is characterized by a high (resp. low) bid-ask-spread Δ_t as well as a high (resp. low) price impact r . Another adaptation that could be considered is to include exogenous parameters in order to provide

indications on the future evolution of the center of the bid-ask-spread. This cannot be achieved easily using SDDP. For the VFA approach, a broad range of adaptations can be considered because the algorithm is able to account for it as long as it is present in the data and in the value function parametrization⁷. Some of the parameters that we would be interested in including in the parametrization are (a) the renewable forecasts and last realizations; (b) the load forecasts and last realizations; (c) the last realizations of the imbalance price; (d) a representation of the trend in the CIM price; (e) indicators about the bid-ask-spread.

In conclusion, both SDDP and VFA present relative strengths and weaknesses. SDDP is empirically observed to be faster and can accommodate more realistic features of the CIM with fewer efforts. On the other hand, VFA seems more flexible in including exogenous parameters that can explain price evolution.

⁷Notice that including these parameters properly in the value function parametrization remains a challenging problem.

3.6 Appendix A: Computation of the Value Function

3.6.1 Proof of Proposition 1

We know that the value function for the j^{th} option at times step $t + 1$ is given by:

$$V_{t+1,j}(s_{t+1}, p_{t+1}) = \max_{q_{t+1} \in Y_j} p_{t+1}q_{t+1} - \Delta_{t+1}q_{t+1} - r(q_{t+1})^2 + V_j^*(s_{t+1} - q_{t+1}, p_{t+1})$$

By developing $V_j^*(s_{t+1} - q_{t+1}, p_{t+1})$ in this expression, we obtain:

$$\begin{aligned} V_{t+1,j}(s_{t+1}, p_{t+1}) &= \max_{q_{t+1} \in Y_j} p_{t+1}q_{t+1} - \Delta_{t+1}q_{t+1} - r(q_{t+1})^2 + (s_{t+1} - q_{t+1})p_{t+1} \\ &\quad - \frac{s_{t+1} - q_{t+1}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{(s_{t+1} - q_{t+1})^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\ &= \max_{q_{t+1} \in Y_j} -\Delta_{t+1}q_{t+1} - r(q_{t+1})^2 + p_{t+1}s_{t+1} - \frac{s_{t+1} - q_{t+1}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) \\ &\quad - r \frac{(s_{t+1} - q_{t+1})^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \end{aligned}$$

We compute the point at which the gradient of the expression vanishes in order to obtain the unconstrained optimal quantity to trade q_{t+1}^* .

$$\begin{aligned} 0 &= -\Delta_{t+1} - 2rq_{t+1}^* + \left(\sum_{i=0}^j \frac{\Delta_i}{j+1} \right) - 2r \frac{(s_{t+1} - q_{t+1}^*)}{j+1} \frac{-1}{j+1} \\ 0 &= -\Delta_{t+1} - 2rq_{t+1}^* + \left(\sum_{i=0}^j \frac{\Delta_i}{j+1} \right) - \frac{2rq_{t+1}^*}{j+1} + \frac{2rs_{t+1}}{j+1} \\ 0 &= -\Delta_{t+1} - 2rq_{t+1}^* \left(\frac{j+1+1}{j+1} \right) + \left(\sum_{i=0}^j \frac{\Delta_i}{j+1} \right) + \frac{2rs_{t+1}}{j+1} \\ 2rq_{t+1}^* \left(\frac{j+2}{j+1} \right) &= -\Delta_{t+1} + \left(\sum_{i=0}^j \frac{\Delta_i}{j+1} \right) + \frac{2rs_{t+1}}{j+1} \\ q_{t+1}^* &= \frac{s_{t+1}}{j+2} - \frac{(j+1)\Delta_{t+1} - \left(\sum_{i=0}^j \Delta_i \right)}{2r(j+2)} \\ &= \frac{s_{t+1}}{j+2} - C_{t+1,j} \end{aligned}$$

where we define $C_{t+1,j} = \frac{(j+1)\Delta_{t+1} - \left(\sum_{i=0}^j \Delta_i \right)}{2r(j+2)}$

3.6.2 Proof of Proposition 2

We evaluate the value function of option j at the optimal unconstrained solution.

$$\begin{aligned}
V_{t+1,j}^u(s_{t+1}, p_{t+1}) &= -\Delta_{t+1}q_{t+1}^* - r(q_{t+1}^*)^2 + p_{t+1}s_{t+1} - \frac{s_{t+1} - q_{t+1}^*}{j+1} \left(\sum_{i=0}^j \Delta_i \right) \\
&\quad - r \frac{(s_{t+1} - q_{t+1}^*)^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&= -\Delta_{t+1} \left(\frac{s_{t+1}}{j+2} - C_{t+1,j} \right) - r \left(\frac{s_{t+1}}{j+2} - C_{t+1,j} \right)^2 + p_{t+1}s_{t+1} \\
&\quad - \frac{s_{t+1} - \left(\frac{s_{t+1}}{j+2} - C_{t+1,j} \right)}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{\left(s_{t+1} - \left(\frac{s_{t+1}}{j+2} - C_{t+1,j} \right) \right)^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&= -\Delta_{t+1} \frac{s_{t+1}}{j+2} + \Delta_{t+1} C_{t+1,j} - r \frac{s_{t+1}^2}{(j+2)^2} - r C_{t+1,j}^2 + 2r \frac{s_{t+1}}{j+2} C_{t+1,j} + p_{t+1}s_{t+1} \\
&\quad - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - \frac{C_{t+1,j}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{\left(\frac{(j+1)s_{t+1}}{j+2} + C_{t+1,j} \right)^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&= p_{t+1}s_{t+1} - \Delta_{t+1} \frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) + \Delta_{t+1} C_{t+1,j} - r \frac{s_{t+1}^2}{(j+2)^2} - r C_{t+1,j}^2 \\
&\quad + 2r \frac{s_{t+1}}{j+2} C_{t+1,j} - \frac{C_{t+1,j}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{(j+1)s_{t+1}^2}{(j+2)^2} - \frac{r}{j+1} C_{t+1,j}^2 \\
&\quad - \frac{2r}{j+2} s_{t+1} C_{t+1,j} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&= p_{t+1}s_{t+1} - \Delta_{t+1} \frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \Delta_{t+1} C_{t+1,j} - r C_{t+1,j}^2 \\
&\quad - \frac{C_{t+1,j}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - \frac{r}{j+1} C_{t+1,j}^2 + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&= p_{t+1}s_{t+1} - \Delta_{t+1} \frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&\quad + C_{t+1,j} \left(\Delta_{t+1} - r C_{t+1,j} - \left(\sum_{i=0}^j \frac{\Delta_i}{j+1} \right) - \frac{r}{j+1} C_{t+1,j} \right)
\end{aligned}$$

$$\begin{aligned}
&= p_{t+1}s_{t+1} - \Delta_{t+1}\frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&+ C_{t+1,j} \left(\Delta_{t+1} - \frac{r(j+2)C_{t+1,j}}{j+1} - \left(\sum_{i=0}^j \frac{\Delta_i}{j+1} \right) \right) \\
&= p_{t+1}s_{t+1} - \Delta_{t+1}\frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&+ C_{t+1,j} \left(\Delta_{t+1} - \frac{r(j+2) \left(\frac{(j+1)\Delta_{t+1} - (\sum_{i=0}^j \Delta_i)}{2r(j+2)} \right)}{j+1} - \left(\sum_{i=0}^j \frac{\Delta_i}{j+1} \right) \right) \\
&= p_{t+1}s_{t+1} - \Delta_{t+1}\frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&+ C_{t+1,j} \left(\Delta_{t+1} - \frac{\left((j+1)\Delta_{t+1} - (\sum_{i=0}^j \Delta_i) \right)}{2(j+1)} - \left(\sum_{i=0}^j \frac{\Delta_i}{j+1} \right) \right) \\
&= p_{t+1}s_{t+1} - \Delta_{t+1}\frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&+ C_{t+1,j} \left(\frac{\left((j+1)\Delta_{t+1} - (\sum_{i=0}^j \Delta_i) \right)}{2(j+1)} \right) \\
&= p_{t+1}s_{t+1} - \Delta_{t+1}\frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&+ r \frac{j+2}{j+1} C_{t+1,j} \left(\frac{\left((j+1)\Delta_{t+1} - (\sum_{i=0}^j \Delta_i) \right)}{2r(j+2)} \right) \\
&= p_{t+1}s_{t+1} - \Delta_{t+1}\frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 + r \frac{j+2}{j+1} C_{t+1,j}^2
\end{aligned}$$

3.6.3 Proof of Proposition 3

We evaluate the value function of option j at $q_{t+1} = 0$.

$$\begin{aligned}
V_{t+1,j}^0 &= -\Delta_{t+1}q_{t+1} - r(q_{t+1})^2 + p_{t+1}s_{t+1} - \frac{s_{t+1} - q_{t+1}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) \\
&\quad - r \frac{(s_{t+1} - q_{t+1})^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&= p_{t+1}s_{t+1} - \frac{s_{t+1}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2
\end{aligned}$$

3.7 Appendix B: Comparison of the Value Functions

3.7.1 Proof of Proposition 4

We compare the value function, evaluated at the unconstrained optimal solution for two consecutive options $j-1$ and j . We show that the one corresponding to the highest option (j) is always higher than the other one ($j-1$).

$$V_{t+1,j}^u(s_{t+1}, p_{t+1}) - V_{t+1,j-1}^u(s_{t+1}, p_{t+1}) = D_1 - D_2$$

where

$$\begin{aligned}
D_1 &= p_{t+1}s_{t+1} - \Delta_{t+1} \frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} \\
&\quad - p_{t+1}s_{t+1} + \Delta_{t+1} \frac{s_{t+1}}{j+1} + \frac{s_{t+1}}{j+1} \left(\sum_{i=0}^{j-1} \Delta_i \right) + r \frac{s_{t+1}^2}{(j+1)} \\
D_2 &= \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 + r \frac{j+2}{j+1} C_{t+1,j}^2 - \sum_{i=1}^{j-1} \frac{(i+1)r}{i} C_i^2 - r \frac{j+1}{j} C_{t+1,j-1}^2
\end{aligned}$$

In order to simplify the development, we develop D_1 and D_2 separately.

$$\begin{aligned}
D_1 &= p_{t+1}s_{t+1} - \Delta_{t+1} \frac{s_{t+1}}{j+2} - \frac{s_{t+1}}{j+2} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{(j+2)} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&\quad + r \frac{j+2}{j+1} C_{t+1,j}^2 - p_{t+1}s_{t+1} + \Delta_{t+1} \frac{s_{t+1}}{j+1} + \frac{s_{t+1}}{j+1} \left(\sum_{i=0}^{j-1} \Delta_i \right) + r \frac{s_{t+1}^2}{(j+1)} \\
&\quad - \sum_{i=1}^{j-1} \frac{(i+1)r}{i} C_i^2 - r \frac{j+1}{j} C_{t+1,j-1}^2 \\
&= \frac{s_{t+1}}{(j+1)(j+2)} \Delta_{t+1} + \frac{s_{t+1}}{(j+1)(j+2)} \left((j+2) \left(\sum_{i=0}^{j-1} \Delta_i \right) - (j+1) \left(\sum_{i=0}^j \Delta_i \right) \right) \\
&\quad + r \frac{s_{t+1}^2}{(j+1)(j+2)} \\
&= \frac{s_{t+1}}{(j+1)(j+2)} \Delta_{t+1} + \frac{s_{t+1}}{(j+1)(j+2)} \left(\left(\sum_{i=0}^{j-1} \Delta_i \right) - (j+1) \Delta_j \right) + r \frac{s_{t+1}^2}{(j+1)(j+2)} \\
&= \frac{s_{t+1}}{(j+1)(j+2)} \Delta_{t+1} + \frac{s_{t+1}}{(j+1)(j+2)} \left(\sum_{i=0}^{j-1} \Delta_i \right) - \frac{s_{t+1}}{(j+2)} \Delta_j + r \frac{s_{t+1}^2}{(j+1)(j+2)}
\end{aligned}$$

$$\begin{aligned}
D_2 &= \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 + r \frac{j+2}{j+1} C_{t+1,j}^2 - \sum_{i=1}^{j-1} \frac{(i+1)r}{i} C_i^2 - r \frac{j+1}{j} C_{t+1,j-1}^2 \\
&= \frac{(j+1)r}{j} C_j^2 + r \frac{j+2}{j+1} C_{t+1,j}^2 - r \frac{j+1}{j} C_{t+1,j-1}^2 \\
&= \frac{(j+1)r}{j} \left(\frac{j\Delta_j - \left(\sum_{i=0}^{j-1} \Delta_i \right)^2}{2r(j+1)} \right)^2 + r \frac{j+2}{j+1} \left(\frac{(j+1)\Delta_{t+1} - \left(\sum_{i=0}^j \Delta_i \right)^2}{2r(j+2)} \right)^2 \\
&\quad - r \frac{j+1}{j} \left(\frac{j\Delta_{t+1} - \left(\sum_{i=0}^{j-1} \Delta_i \right)^2}{2r(j+1)} \right)^2 \\
&= \frac{(j+1)r}{j} \frac{\left(j^2\Delta_j^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_j \left(\sum_{i=0}^{j-1} \Delta_i \right) \right)}{4r^2(j+1)^2} \\
&\quad + \frac{(j+2)r}{j+1} \frac{\left((j+1)^2\Delta_{t+1}^2 + \left(\sum_{i=0}^j \Delta_i \right)^2 - 2(j+1)\Delta_{t+1} \left(\sum_{i=0}^j \Delta_i \right) \right)}{4r^2(j+2)^2} \\
&\quad - \frac{(j+1)r}{j} \frac{\left(j^2\Delta_{t+1}^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_{t+1} \left(\sum_{i=0}^{j-1} \Delta_i \right) \right)}{4r^2(j+1)^2}
\end{aligned}$$

$$\begin{aligned}
&= \frac{j^2 \Delta_j^2 + \left(\sum_{i=0}^{j-1} \Delta_i\right)^2 - 2j \Delta_j \left(\sum_{i=0}^{j-1} \Delta_i\right)}{4rj(j+1)} \\
&- \frac{j^2 \Delta_{t+1}^2 + \left(\sum_{i=0}^{j-1} \Delta_i\right)^2 - 2j \Delta_{t+1} \left(\sum_{i=0}^{j-1} \Delta_i\right)}{4rj(j+1)} \\
&+ \frac{(j+1)^2 \Delta_{t+1}^2 + \left(\sum_{i=0}^j \Delta_i\right)^2 - 2(j+1) \Delta_{t+1} \left(\sum_{i=0}^j \Delta_i\right)}{4r(j+1)(j+2)} \\
&= \frac{j^2 (\Delta_j^2 - \Delta_{t+1}^2) - 2j (\Delta_j - \Delta_{t+1}) \left(\sum_{i=0}^{j-1} \Delta_i\right)}{4rj(j+1)} \\
&+ \frac{(j+1)^2 \Delta_{t+1}^2 + \left(\sum_{i=0}^j \Delta_i\right)^2 - 2(j+1) \Delta_{t+1} \left(\sum_{i=0}^j \Delta_i\right)}{4r(j+1)(j+2)} \\
&= \frac{j(\Delta_j^2 - \Delta_{t+1}^2) - 2(\Delta_j - \Delta_{t+1}) \left(\sum_{i=0}^{j-1} \Delta_i\right)}{4r(j+1)} \\
&+ \frac{(j+1)^2 \Delta_{t+1}^2 + \left(\sum_{i=0}^j \Delta_i\right)^2 - 2(j+1) \Delta_{t+1} \left(\sum_{i=0}^j \Delta_i\right)}{4r(j+1)(j+2)} \\
&= \frac{j(j+2)(\Delta_j^2 - \Delta_{t+1}^2) - 2(j+2)(\Delta_j - \Delta_{t+1}) \left(\sum_{i=0}^{j-1} \Delta_i\right)}{4r(j+1)(j+2)} \\
&+ \frac{(j+1)^2 \Delta_{t+1}^2 + \left(\sum_{i=0}^j \Delta_i\right)^2 - 2(j+1) \Delta_{t+1} \left(\sum_{i=0}^j \Delta_i\right)}{4r(j+1)(j+2)} \\
&= \frac{j(j+2) \Delta_j^2 + \Delta_{t+1}^2 - (2(j+2) \Delta_j - 2 \Delta_{t+1}) \left(\sum_{i=0}^{j-1} \Delta_i\right) - 2(j+1) \Delta_j \Delta_{t+1}}{4r(j+1)(j+2)} \\
&+ \frac{\left(\sum_{i=0}^j \Delta_i\right)^2}{4r(j+1)(j+2)} \\
&= \frac{j(j+2) \Delta_j^2 + \Delta_{t+1}^2 - 2(j+2) \Delta_j \left(\sum_{i=0}^{j-1} \Delta_i\right) + 2 \Delta_{t+1} \left(\sum_{i=0}^{j-1} \Delta_i\right) - 2(j+1) \Delta_j \Delta_{t+1}}{4r(j+1)(j+2)} \\
&+ \frac{\left(\sum_{i=0}^j \Delta_i\right)^2}{4r(j+1)(j+2)} \\
&= \frac{j(j+2) \Delta_j^2 + \Delta_{t+1}^2 - 2(j+2) \Delta_j \left(\sum_{i=0}^{j-1} \Delta_i\right) + 2 \Delta_{t+1} \left(\sum_{i=0}^{j-1} \Delta_i\right) - 2(j+1) \Delta_j \Delta_{t+1}}{4r(j+1)(j+2)} \\
&+ \frac{\left(\sum_{i=0}^{j-1} \Delta_i\right)^2 + 2 \Delta_j \left(\sum_{i=0}^{j-1} \Delta_i\right) + \Delta_j^2}{4r(j+1)(j+2)}
\end{aligned}$$

$$\begin{aligned}
&= \frac{(j+1)^2 \Delta_j^2 + \Delta_{t+1}^2 - 2(j+1)\Delta_j \left(\sum_{i=0}^{j-1} \Delta_i \right) + 2\Delta_{t+1} \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r(j+1)(j+2)} \\
&+ \frac{-2(j+1)\Delta_j \Delta_{t+1} + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2}{4r(j+1)(j+2)} \\
&= \frac{r}{(j+1)(j+2)} \left(\frac{\Delta_{t+1} + \sum_{i=0}^{j-1} \Delta_i - (j+1)\Delta_j}{2r} \right)^2
\end{aligned}$$

After computing D_1 and D_2 , we can write the difference of the two value functions.

$$\begin{aligned}
&V_{t+1,j}^u(s_{t+1}, p_{t+1}) - V_{t+1,j-1}^u(s_{t+1}, p_{t+1}) \\
&= \frac{s_{t+1}}{(j+1)(j+2)} \Delta_{t+1} + \frac{s_{t+1}}{(j+1)(j+2)} \left(\sum_{i=0}^{j-1} \Delta_i \right) - \frac{s_{t+1}}{(j+2)} \Delta_j + r \frac{s_{t+1}^2}{(j+1)(j+2)} \\
&+ \frac{r}{(j+1)(j+2)} \left(\frac{\Delta_{t+1} + \sum_{i=0}^{j-1} \Delta_i - (j+1)\Delta_j}{2r} \right)^2 \\
&= \frac{r}{(j+1)(j+2)} \left(s_{t+1} + \frac{\Delta_{t+1} + \sum_{i=0}^{j-1} \Delta_i - (j+1)\Delta_j}{2r} \right)^2
\end{aligned}$$

As this expression is always positive, it means that, if evaluated at the unconstrained optimal solution, the value function of option j is always higher than the value function of option $j-1$ for any s_{t+1} and p_{t+1} .

3.7.2 Proof of Proposition 5

We compute, for which value of s_{t+1} , the value function of option j evaluated at 0, $V_{t+1,j}^0(s_{t+1}, p_{t+1})$ is higher than the value function of option $j-1$ evaluated at the unconstrained optimal solution, $V_{t+1,j-1}^u(s_{t+1}, p_{t+1})$.

$$\begin{aligned}
& 0 \leq V_{t+1,j}^0(s_{t+1}, p_{t+1}) - V_{t+1,j-1}^u(s_{t+1}, p_{t+1}) \\
& \Leftrightarrow 0 \leq p_{t+1}s_{t+1} - \frac{s_{t+1}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 - p_{t+1}s_{t+1} \\
& \quad + \Delta_{t+1} \frac{s_{t+1}}{j+1} + \frac{s_{t+1}}{j+1} \left(\sum_{i=0}^{j-1} \Delta_i \right) + r \frac{s_{t+1}^2}{(j+1)} - \sum_{i=1}^{j-1} \frac{(i+1)r}{i} C_i^2 - r \frac{j+1}{j} C_{t+1,j-1}^2 \\
& \Leftrightarrow 0 \leq (\Delta_{t+1} - \Delta_j) \frac{s_{t+1}}{j+1} + \frac{(j+1)r}{j} C_j^2 - r \frac{j+1}{j} C_{t+1,j-1}^2 \\
& \Leftrightarrow 0 \leq (\Delta_{t+1} - \Delta_j) \frac{s_{t+1}}{j+1} + \frac{(j+1)r}{j} \left(\frac{j\Delta_j - \left(\sum_{i=0}^{j-1} \Delta_i \right)^2}{2r(j+1)} \right)^2 \\
& \quad - r \frac{j+1}{j} \left(\frac{j\Delta_{t+1} - \left(\sum_{i=0}^{j-1} \Delta_i \right)^2}{2r(j+1)} \right)^2 \\
& \Leftrightarrow 0 \leq (\Delta_{t+1} - \Delta_j) \frac{s_{t+1}}{j+1} + \frac{(j+1)r}{j} \left(\frac{j^2\Delta_j^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_j \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r^2(j+1)^2} \right) \\
& \quad - r \frac{j+1}{j} \left(\frac{j^2\Delta_{t+1}^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_{t+1} \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r^2(j+1)^2} \right) \\
& \Leftrightarrow 0 \leq (\Delta_{t+1} - \Delta_j) \frac{s_{t+1}}{j+1} + \left(\frac{j^2\Delta_j^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_j \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4rj(j+1)} \right) \\
& \quad - \left(\frac{j^2\Delta_{t+1}^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_{t+1} \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r(j+1)j} \right) \\
& \Leftrightarrow 0 \leq (\Delta_{t+1} - \Delta_j) \frac{s_{t+1}}{j+1} + \left(\frac{j^2 \left(\Delta_j^2 - \Delta_{t+1}^2 \right) - 2j(\Delta_j - \Delta_{t+1}) \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4rj(j+1)} \right) \\
& \Leftrightarrow 0 \leq (\Delta_{t+1} - \Delta_j) s_{t+1} + \left(\frac{j \left(\Delta_j^2 - \Delta_{t+1}^2 \right) - 2(\Delta_j - \Delta_{t+1}) \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r} \right) \\
& \Leftrightarrow s_{t+1} \geq \left(\frac{j(\Delta_j + \Delta_{t+1}) - 2 \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r} \right)
\end{aligned}$$

3.7.3 Proof of Proposition 6

We compare the value function, evaluated at 0 for two consecutive options $j-1$ and j . We show that the value function corresponding to the highest option (j) is always higher

than the value function of option $(j - 1)$.

$$\begin{aligned}
0 &\leq p_{t+1}s_{t+1} - \frac{s_{t+1}}{j+1} \left(\sum_{i=0}^j \Delta_i \right) - r \frac{s_{t+1}^2}{j+1} + \sum_{i=1}^j \frac{(i+1)r}{i} C_i^2 \\
&\quad - p_{t+1}s_{t+1} + \frac{s_{t+1}}{j} \left(\sum_{i=0}^{j-1} \Delta_i \right) + r \frac{s_{t+1}^2}{j} - \sum_{i=1}^{j-1} \frac{(i+1)r}{i} C_i^2 \\
&\Leftrightarrow 0 \leq \frac{s_{t+1}}{j(j+1)} \left(\sum_{i=0}^{j-1} \Delta_i \right) - \frac{s_{t+1}}{(j+1)} \Delta_j + r \frac{s_{t+1}^2}{j(j+1)} + \frac{(j+1)r}{j} C_j^2 \\
&\Leftrightarrow 0 \leq \frac{s_{t+1}}{j(j+1)} \left(\sum_{i=0}^{j-1} \Delta_i \right) - \frac{s_{t+1}}{(j+1)} \Delta_j + r \frac{s_{t+1}^2}{j(j+1)} + \frac{(j+1)r}{j} \left(\frac{j\Delta_j - \left(\sum_{i=0}^{j-1} \Delta_i \right)}{2r(j+1)} \right)^2 \\
&\Leftrightarrow 0 \leq \frac{s_{t+1}}{j(j+1)} \left(\sum_{i=0}^{j-1} \Delta_i \right) - \frac{s_{t+1}}{(j+1)} \Delta_j + r \frac{s_{t+1}^2}{j(j+1)} \\
&\quad + \frac{(j+1)r}{j} \left(\frac{j^2\Delta_j^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_j \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r^2(j+1)^2} \right) \\
&\Leftrightarrow 0 \leq \frac{s_{t+1}}{j(j+1)} \left(\sum_{i=0}^{j-1} \Delta_i \right) - \frac{s_{t+1}}{(j+1)} \Delta_j + r \frac{s_{t+1}^2}{j(j+1)} \\
&\quad + \left(\frac{j^2\Delta_j^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_j \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4rj(j+1)} \right) \\
&\Leftrightarrow 0 \leq s_{t+1} \left(\sum_{i=0}^{j-1} \Delta_i \right) - js_{t+1}\Delta_j + rs_{t+1}^2 + \left(\frac{j^2\Delta_j^2 + \left(\sum_{i=0}^{j-1} \Delta_i \right)^2 - 2j\Delta_j \left(\sum_{i=0}^{j-1} \Delta_i \right)}{4r} \right) \\
&\Leftrightarrow 0 \leq r \left(s_{t+1} + \frac{\left(\sum_{i=0}^{j-1} \Delta_i \right) - j\Delta_j}{2r} \right)^2
\end{aligned}$$

As this expression is always positive, it means that, if evaluated at 0, the value function of option j is always higher than the value function of option $j - 1$ for any s_{t+1} and p_{t+1} .

Chapter 4

Market Design Options for Scarcity Pricing in European Balancing Markets

4.1 Introduction

4.1.1 Comparison between European and US markets

Scarcity pricing based on operating reserve demand curves has been initially proposed by Stoft [Sto02]. This concept has been further explored by Hogan in the context of US-type markets [Hog05, Hog13]. It is therefore worth mentioning some major axes of differences between the European and US-type markets.

- **Co-optimization of balancing capacity and energy:** In Europe, the day-ahead market and the balancing capacity market are cleared independently (section 1.3.1 and 1.3.4.1). On the contrary, in US markets, there is a co-optimization of reserve and energy.
- **Unique price signal for real-time energy:** In Europe, BRPs are settled for their real-time energy deviations at the imbalance price (section 1.3.5), whereas BSPs are settled for their real-time deviations at the balancing price (section 1.3.4.2). On the contrary, in US-style markets, real-time energy is typically cleared at the same price.
- **Real-time market for reserve capacity:** In the US standard market design, there is a market for real-time reserve capacity which is not the case in Europe. A real-time market for reserve capacity settles reserve imbalances at a real-time reserve price. Putting such a market in place in an EU design would imply that (i) *free* bids¹ that are available in real time are paid even if they have not sold balancing capacity in the day ahead, and (ii) resources that are activated as upward balancing energy and are no longer available as balancing capacity in real time are required to buy back their day-ahead balancing capacity at the real-time price of balancing capacity.

These differences in design create challenges in the valuation of reserve, as we discuss next.

¹Free bids refer to bids that are not linked to a capacity cleared in the day-ahead balancing capacity market.

4.1.2 Motivation

The accurate valuation of energy and reserve is an increasingly crucial function of real-time markets in a regime of large-scale renewable energy integration. Operating reserve demand curves (ORDCs) [Hog05] have been proposed as a means for achieving this important goal. ORDC adders are computed on the basis of available reserve in the system as presented in section 1.3.6.

ORDC adders have been adopted in Texas [ERC15], and their adoption is moving forward in PJM [HP19]. The Electricity Balancing Guideline of the European Commission, which is the reference text for European balancing legislation (and which we will refer to as “EBGL” hereafter), introduces the legal possibility of implementing ORDC adders by referring to the mechanism as a “scarcity pricing function” in article 44(3) of the legislation [Eur17].

Belgium has made steps in advancing the implementation of scarcity pricing. A series of preliminary analyses have focused on quantifying the possible implications of the mechanism for resources that can provide balancing capacity to the system [PS17, PSB18]. The Belgian regulator and TSO [ELI18] have collaborated towards computing and publishing scarcity adders based on the “available reserve capacity” (ARC) of the system. These adders are computed for every quarter of the day, and published one day after operations. The next step of the process is to decide the details of the integration of ORDC in the Belgium market. The target is to have a market design that sends a stable investment signal to investors. This is achieved by the back-propagation of the balancing capacity price to forward markets. Therefore, there is a need for a methodology which is able to detect the ability of the different proposed designs to back-propagate this balancing capacity value to forward markets.

4.1.3 Existing Modeling Frameworks

In previous analysis [PSd19], stochastic equilibrium has been used for representing the back-propagation effect quantitatively. The stochastic equilibrium framework that has been developed, which has originally been applied in the context of investment [RS15, ES11], reveals the strengths and weaknesses of different market design choices in back-propagating the value of balancing capacity to forward balancing capacity auctions. However, the stochastic equilibrium framework encountered an immediate weakness from the outset during discussions with stakeholders: it embeds the law of one price [Cra17], meaning that the model assumes a unique market for real-time energy, and therefore a unique price for real-time energy. This assumption contradicts the practice of using imbalance prices for BRP settlement that are different from balancing prices for BSP settlement². To put it differently: whereas stochastic equilibrium can be used for understanding the effect of certain market design choices on the back-propagation of balancing capacity prices to forward markets, it cannot be used for assessing the validity of different mixtures of BSP and BRP settlement on this back-propagation.

An alternative model that is developed in this chapter is the representation of the balancing market as a Markov Decision Process (MDP). Our approach is inspired by a

²Note that this assumption is also not compatible with the future implementation of the European balancing platforms (MARI and PICASSO). Indeed, the price generated by these separate platforms for the activated energy of the different reserve products can be different. The reason is that, as stated in ACER decision [ACE20], the price for each reserve product should be set at the marginal price of each platform and cannot be set at the marginal price between the platforms.

growing body of work on the application of agent-based models to the analysis of electricity markets. In early work on this topic, Bunn and coauthors [BB01,BO01] analyze the effect of a change of design in the England and Wales market. In recent work, with the broader use of Reinforcement Learning techniques such as Q-Learning [WD92], researchers have applied MDPs [YLP10,ND07] in more complex settings. However, these classical Reinforcement Learning techniques are inefficient for high-dimensional problems because they rely on the discretization of the state and action space (see section 1.4.3). This problem has been overcome recently by the development of deep-learning [YQLS19,YQS⁺20]. As we discuss in section 4.2, our problem is low-dimensional, and therefore we rely on the standard Q-learning algorithm [WD92].

In the context of our analysis, we consider BRPs and BSPs as agents that engage in trade in a balancing market, and develop trading strategies *given* different market design options. We then test the ability of agents to infer the value of the balancing capacity that they offer to the market under different market design choices, and thus the ability of different market design choices to back-propagate the value of this reserve capacity in forward reserve markets.

The MDP framework offers powerful modeling flexibility. However, since one is limited to observing the outcome of a simulation, it might be difficult to extract conclusions about the role of a market design in driving a certain outcome. For this reason, we supplement our MDP-based market simulation framework with an analytical characterization of the best response of market agents to different balancing market design choices under an assumption of perfect competition. The MDP simulation framework is then used for providing tangible evidence for the behavior that the analytical mathematical framework predicts, which can be valuable for discussions with stakeholders.

By comparison, the stochastic equilibrium approach [PSdMd20] combines the advantages of analytical insights and numerical scalability in a single modeling framework. Concretely, the complementarity conditions of the stochastic equilibrium model provide generalizable conclusions about the effect of market design choices on the back-propagation of balancing capacity prices (see, for instance, the discussion in page 21 of [PSdMd20]). Assuming risk-neutral market agents, the stochastic equilibrium models of [PSdMd20] can further be expressed as equivalent tractable two-stage or three-stage stochastic programming optimization problems. However, it is not clear how the stochastic equilibrium framework can be adapted in order to account for how agents internalize opportunity costs in their *bidding* behavior, and for the fact that the EU market design allows BSPs and BRPs to trade at different settlement prices.

4.1.4 Contributions and Structure

Our claimed contribution in this chapter is twofold. We propose an analytical framework for analyzing European balancing markets which we supplement by an MDP-based market simulator. And we use our framework to arrive at concrete insights and recommendations regarding the design of the European balancing market. One important recommendation is to introduce a real-time balancing capacity market in the European balancing design.

The remainder of the chapter is structured as follows. In section 4.2, we describe the basic MDP framework for simulating the balancing market. In section 4.3, we present various market design options, show their specificity related to the European balancing market and describe the adaptations that need to be applied to the basic MDP to represent them. After that, in section 4.4, we analyze these different market design options under

an assumption of perfect competition, and summarize our main conclusions regarding the strengths and weaknesses of different market design proposals. In section 4.5 we validate our theoretical results by applying the MDP simulation framework of section 4.2.2 in order to test the ability of different balancing market design options in back-propagating the value of reserve capacity to forward reserve markets. We conclude our analysis and discuss prospects for future research in section 4.6.

4.2 Modeling the Balancing Market in the MDP Framework

4.2.1 Problem Description

As we mention in section 1.3.4, each BSP must be attributed to a unique BRP according to article 18(4).d of the EBGL [Eur17]. Virtual traders³ are not present in our model because they are not legal in European balancing markets. Indeed, offers in the balancing energy market have to be backed by assets that have been pre-qualified by the TSO [ELIc]. The TSO verifies regularly the ability of BSP assets to provide the promised balancing energy. Moreover, virtual trading of reserve is also not foreseen⁴. Without loss of generality, therefore, we consider a generic agent participating in the balancing market as one which owns (i) a pool of uncontrollable assets that impose a price-inelastic imbalance (positive or negative) to the system as well as (ii) a set of controllable assets with marginal cost C that is private⁵ information of the agent, and with a total upward capacity P^+ and downward capacity P^- that is common knowledge for the TSO. The controllable set of assets can be offered to the balancing market. The sequence of events faced by the agent is illustrated in Fig. 4.1. The agent starts by placing a bid in the day-ahead balancing capacity at a price p^R and a quantity q^R . The TSO collects the bids of every agents, clears the market using an ORDC and announces the day-ahead balancing capacity price $\lambda^{R,DA}$ as well as the quantity cleared for each agent qa^R . Then, the agent can submit a bid to the balancing market with a price p and quantity q where q should be at least equal to the quantity cleared in the day-ahead balancing capacity market qa^R . After that, the agent observes the imbalance of the part of its portfolio with uncontrollable assets Imb . The convention is that positive imbalance means that a portfolio is consuming more than it is producing. With this information, the agent can decide if it wants to resort to active imbalance⁶ with its leftover capacity⁷. Finally, the TSO (i) observes the system imbalance (ii) activates the necessary balancing capacities to cover it (iii) computes the balancing price λ^B as well as the imbalance price λ^I and realizes the payment to the agents accordingly.

³We refer to virtual traders as market participants that would attempt to arbitrage between the day-ahead and real-time balancing capacity market without physical assets.

⁴Virtual trading of reserve is also not typically foreseen in US markets. In Europe, there is anyway no real-time market for reserve.

⁵Our analysis if the marginal cost of the agent would be known by the other agents or by the TSO would not change.

⁶This means that the agent changes its production, on purpose, compared to the scheduled one without notifying the market. Notice that the representation of this third step is a simplification. In reality, it is a continuous interaction between the agent observing the system imbalance and reacting by performing active imbalance which changes the system imbalance.

⁷The capacity that has not been promised in the balancing energy market.

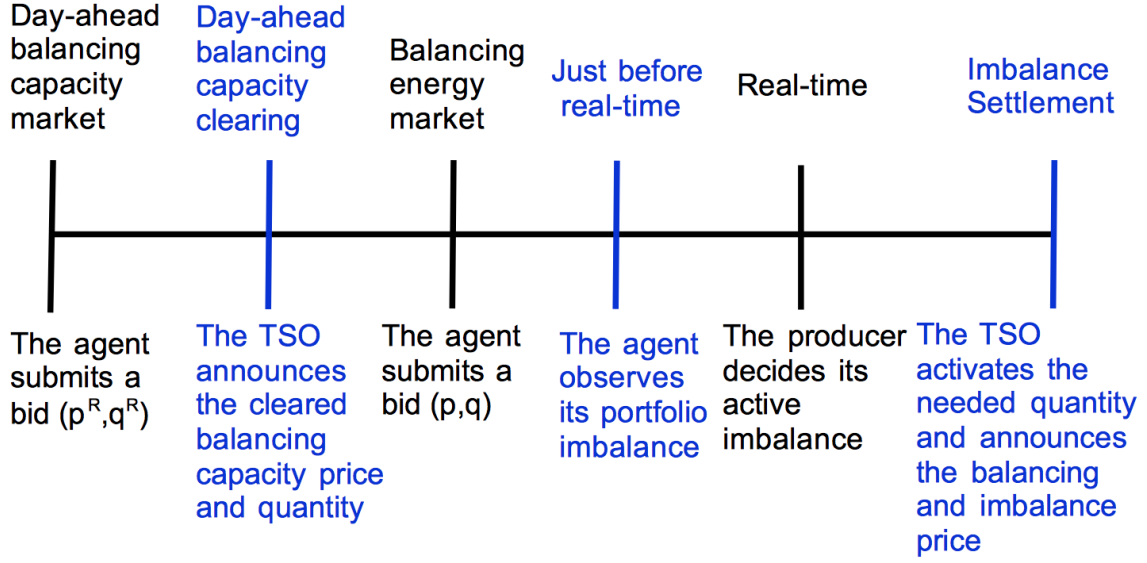


Figure 4.1: Sequence of events faced by an agent in the balancing markets.

4.2.2 MDP Modelling

The decision process faced by the agent can be described as an MDP. It should be noticed that the MDP might be slightly different from one design to another. Therefore, in this section, we present the basis that is common for every design and we present the differences related to each of them in section 4.3.

- Stage 1
 - State: a single element, the default state of the world
 - Action: (p^R, q^R) , the price-quantity offers in the balancing capacity auction
 - Reward: the payment from the balancing capacity auction, $\lambda^{R,DA} \cdot qa^R$, where $\lambda^{R,DA}$ is the forward price of balancing capacity and qa^R is the amount of balancing capacity that has been cleared in the forward balancing capacity auction
- Stage 2
 - State: the capacity qa^R awarded in the balancing capacity auction
 - Action: (p, q) , the price-quantity offers in the balancing platform. The offered quantity can be no less than what has been cleared in the balancing capacity auction, i.e. $q \geq qa^R$.
 - No reward is collected at this stage
- Stage 3:
 - State: (i) the bid price p , (ii) the leftover BSP capacity⁸ $P^+ - q$ after the capacity

⁸Note that the bid quantity q is also implied by the second state. Effectively, the actions of the second stage are states for the third stage. This is required, because the part of the reward of the third stage that depends on BSP payments is a result of these second-stage decisions, and since we need to obey the structure of an MDP, it is necessary to augment the state space in order to maintain the Markov property.

q has been offered to the balancing auction, and (iii) the level of imbalance Imb of an agent.

- Action: How much of the imbalance Imb to cover (this action, denoted as ai and referred to as “active imbalance”, must be limited to the leftover capacity that the BSP has not allocated to the balancing capacity auction, i.e. $ai \leq P^+ - q$).
- Reward: (i) BSP payment for upward/downward activation, expressed as $\lambda^B \cdot qa$, (ii) BRP payment for imbalance settlement⁹, expressed as $-\lambda^I \cdot qi = -\lambda^I \cdot (Imb - ai)$, and (iii) fuel costs related to self-balancing and BSP activation, expressed as $-C \cdot (ai + qa)$.

4.3 Considered Designs

Our analysis will focus on six different market design options. These options are inspired by discussions with stakeholders about different ways in which the European balancing market could be organized so as to more accurately reflect the value of balancing capacity. For each of these designs, we first summarize the design and present the underlying intuition. We also present the adaptations that need to be effected on the vanilla MDP.

4.3.1 Design 1

The default European design is the one described in section 4.2.1 for which the imbalance price is equal to the balancing price, $\lambda^I = \lambda^B$. Note that it has already been argued, using a stochastic equilibrium framework in [PSd19], that this design is unable to generate a forward reserve price signal that reflects the value of reserve capacity. The MDP is exactly the one described in section 4.2.2.

4.3.2 Design 2

For the second design, a difference between the balancing price and the imbalance price is introduced in the model. This is the current practice, for example, in Belgium, where the system operator computes the imbalance price by applying a surcharge α^+ whenever the system is short, or a discount α^- whenever the system is long [ELI19]. Mathematically, the imbalance price in this setting can be expressed as:

$$\lambda^I = \lambda^B + \alpha \quad (4.1)$$

$$\alpha \triangleq \alpha^U \cdot \mathbb{I}[Imb^t > UI] - \alpha^D \cdot \mathbb{I}[Imb^t < LI] \quad (4.2)$$

Here, Imb^t corresponds to the total imbalance of the system. The parameters UI and LI represent the upper and lower imbalance thresholds at which the surcharge or discount apply, respectively.

We consider this design because, as a response to the request of the European Commission for planned market reforms in order to implement scarcity pricing (article 20(3) of regulation 2019/943 [Eur19]) the Belgian government [ELIb] mentions that the imbalance penalty α of Eq. (4.1), “already exhibits quite some characteristics of a scarcity pricing mechanism” [ELIb].

⁹In this section, we will use a generic λ^I for the imbalance price. The definition of this λ^I will differ from design to design and will be described in section 4.3

It is important to note that design (*D2*) relies on imbalance penalties α which depend on the level of system imbalance, which is not to be confused with the level of scarcity in the system. To clarify: a system that is exhibiting a very large positive imbalance is not experiencing scarcity if it carries abundant reserve capacity at the moment in time when the large imbalance occurs.

In practice, the imbalance penalty in Eq. (4.1) depends on the imbalance of the current and previous interval (see Eq. (4.4) below). Therefore, the MDP model that we develop for design (*D2*) requires an additional state variable, the imbalance of the previous balancing interval, which is added to the state vector of stages 2 and 3. It is not added to stage 1 of the problem because this quantity is unknown in day-ahead.

4.3.3 Design 3

Scarcity pricing, as presented in section 1.3.6, introduces a real-time price for balancing capacity, or ORDC adder, which is a function of the instantaneous amount of leftover capacity in the system, λ^R . The question is where this adder should be applied. It has been proposed [Gie19] to apply this adder as an imbalance charge, as an alternative to the α penalty of Eq. (4.1).

$$\lambda^I = \lambda^B + \lambda^R$$

The MDP model that we develop for design (*D3*) is exactly the same as the one for design (*D1*) in section 4.2.2.

4.3.4 Design 4

The intuition for this design is based on the following economical principles:

- **Economic principle 1: law of one price [Cra17].** Real-time energy is a unique product, therefore the buyer and seller should exchange it at the same price.
- **Economic principle 2: back-propagation.** If we put in place a real-time market for balancing capacity, then agents will only sell balancing capacity in forward markets at the value that they would need to buy it back in real time. This second principle is especially crucial, since it allows the value of balancing capacity to back-propagate into forward balancing capacity auctions, and send the signal to investors that the market can support investments in balancing capacity.

These two economical principles translate to the following market design proposals for implementing scarcity pricing in the EU market design [PSd19]:

- **Market design proposal 1:** the introduction of a scarcity adder to the imbalance price.
- **Market design proposal 2:** the application of the same adder to the balancing energy price.
- **Market design proposal 3:** the implementation of an EU real-time market for balancing capacity (equivalently, a market for “balancing capacity imbalances”, in the same way that we operate a market for energy imbalances), which is a missing market in the existing EU balancing design.

In terms of the MDP model, this implies adding λ^R to the balancing price λ^B and to the imbalance price λ^I . Moreover, we need to introduce the following term in settlement:

$$-\lambda^R \cdot qa^R + \lambda^R \cdot (P^+ - qa - ai).$$

This term effectively implies that agents buy back their day-ahead balancing capacity at real-time balancing capacity prices, and sell their entire real-time balancing capacity at real-time balancing capacity prices. Introducing this settlement of real-time balancing capacity imbalances induces agents to bid their balancing capacity in forward markets in a way that anticipates the expected price at which they would be required to buy that balancing capacity back in real time. This effect results in the back-propagation of the scarcity signal.

The mechanism amounts to introducing a real-time market for balancing capacity, and is exactly analogous to the practice of settling *energy* imbalances at prevailing real-time energy prices. Furthermore, the approach is compatible with the EBGL, since one can invoke article 44(3) of the EBGL for introducing a scarcity adder (shortage pricing function in EBGL) to the imbalance price, and article 20(c) of [Eur19] for introducing a scarcity adder on top of the balancing energy price.

The representation of this design requires augmenting the MDP model of section 4.2.2 by adding the awarded day-ahead balancing capacity qa^R to the state of the third time step, since this quantity affects the third-stage payoff under design (D4).

4.3.5 Design 5

Design (D5) is the design proposed by ELIA in its September 2020 report [ELIe]. The aim of this design is to incorporate the scarcity adder of design (D3) into design (D2). Mathematically, the imbalance price can be written as:

- if the system is long, $\lambda^I = \lambda_B - \alpha^D \cdot \mathbb{I}[Imb^t < LI]$
- if the system is short, $\lambda^I = \lambda_B + \max(\alpha^U \cdot \mathbb{I}[Imb^t > UI], \lambda^R)$

The MDP model that we develop for design (D5) is the same as the one for (D2). It requires an additional state variable, the imbalance of the previous balancing interval, which is added to the state vector of stages 2 and 3.

4.3.6 Design 6

We now consider a design that is a mix between what is proposed in (D4) and what ELIA proposes (D5). Specifically, the idea is to keep the real-time market for balancing capacity and the application of the scarcity adder to the balancing energy price presented in (D4) (market design proposal 2 and 3) as well as using the same imbalance settlement as in (D5).

The MDP model that we develop for design (D6) requires (i) one extra variable for stage 2, the imbalance of the previous balancing interval (ii) two extra variables for stage 3, the imbalance of the previous balancing interval and the awarded day-ahead balancing capacity qa^R .

4.4 Analysis under Perfect Competition Assumption

This section analyzes each of the six designs that are introduced in section 4.3 under the simplifying assumption of perfect competition¹⁰. Unveiling difficulties in back-propagating balancing capacity prices in the case of perfect competition suggests fundamental market design problems, and offers insights about what to expect in the simulations of section 4.5.2. Our simplifying assumption can be stated as follows:

Perfect competition assumption: We consider fringe agents, i.e. ones with infinitesimal capacity.

In order to keep the development concise, and since the proof strategy is the same for the different designs, we only present the complete proof for (D1). For the other designs, we present the optimal strategy for the agent. The proofs are available in appendix 4.7.

4.4.1 Generalities

The first step in the proof for all the designs is to demonstrate that there is no loss of generality in considering the case of an agent which has only downward capacity (i.e. $P^+ = 0$ and $P^- < 0$) or the case of an agent which has only upward capacity (i.e. $P^- = 0$ and $P^+ > 0$).

Consider the general case of an agent for which $P^+ > 0$ and $P^- < 0$. Suppose that the agent has offered $q^+ > 0$ of capacity in the balancing auction for upward energy, and $q^- < 0$ of capacity in the balancing auction for downward energy.

The balancing activation payoff is as follows:

$$z_B(\omega) = (\lambda^B(\omega) - C) \cdot (qa^+(q^+, \omega) + qa^-(q^-, \omega))$$

where

- ω corresponds to a realization of system imbalance.
- $qa^+(q^+, \omega)$ is the quantity accepted for upward balancing if the agent bids q^+ in scenario ω .
- $qa^-(q^-, \omega)$ is the quantity accepted for downward balancing if the agent bids q^- in scenario ω .
- $\lambda^B(\omega)$ is the balancing price in scenario ω . Our assumption of focusing on a fringe supplier justifies the assumption of considering $\lambda^B(\omega)$ as not being influenced by the agent decisions q^+ and q^- .

The bid quantities obey the following constraints:

$$\begin{aligned} 0 &\leq q^+ \leq P^+ \\ P^- &\leq q^- \leq 0 \end{aligned}$$

As q^+ and q^- never appear in the same constraint and do not interact in the objective, the balancing payoff z_B is separable in q^+ and q^- . Denote as ai^D the positive active

¹⁰The analytical solutions of design (D1) – (D4) have been originally developed in [Pap20] [PB21]. On the other hand, the proofs of designs (D5) and (D6) are original contributions of this thesis. As the proof strategies are similar, we also reproduce the results for (D1) – (D4) in this thesis.

imbalance (downward regulation), and ai^U as the negative active imbalance (upward regulation). Given a second-stage active imbalance $ai = ai^D - ai^U$ (and an implied imbalance qi), the agent receives an imbalance payoff which is computed as follows:

$$-\lambda^I(\omega) \cdot qi - C \cdot (ai^U - ai^D)$$

with

$$qi = Imb(\omega) + ai^D - ai^U$$

where $Imb(\omega)$ is the imbalance of the agent in scenario ω .

By substituting out the imbalance and considering expectations, the active imbalance optimization is written as:

$$\begin{aligned} z_I = \max_{ai^D, ai^U} & (\mathbb{E}[\lambda^I] - C) \cdot ai^U + (C - \mathbb{E}[\lambda^I]) \cdot ai^D - \mathbb{E}[\lambda^I \cdot Imb] \\ & ai^U + q^+ \leq P^+ \\ & ai^U \leq P^+ \\ & ai^D - q^- \leq -P^- \\ & ai^D \leq -P^- \\ & ai^D, ai^U \geq 0 \end{aligned}$$

Note that the upward active imbalance ai^U only interacts with the upward capacity bid q^+ , and the downward active imbalance ai^D only interacts with the downward bid capacity q^- . Thus, the problem is separable in (ai^U, q^+) and in (ai^D, q^-) , insofar as z_I is concerned. And since the payoff z_B is separable in q^+ and q^- , the desired conclusion follows.

4.4.2 Design 1

For the first design, we have four different situations depending on the agent characteristics (i) agents with upward capacity and for which $\mathbb{E}[\lambda^B] < C$; (ii) agents with upward capacity and for which $\mathbb{E}[\lambda^B] \geq C$; (iii) agents with downward capacity and for which $\mathbb{E}[\lambda^B] \geq C$; and (iv) agents with downward capacity and for which $\mathbb{E}[\lambda^B] < C$.

(D1) Case 1: $C \leq \mathbb{E}[\lambda^B]$, $P^+ > 0$, $P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ \geq 0$ (and therefore $q \geq 0$):

$$\begin{aligned} & \max_{ai} (\mathbb{E}[\lambda^B] - C) \cdot ai - \mathbb{E}[\lambda^B] \cdot Imb \\ & ai + q \leq P^+ \\ & ai \geq 0 \end{aligned}$$

As $C \leq \mathbb{E}[\lambda^B]$, we have $ai^* = P^+ - q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B] - C) \cdot (P^+ - q) + D$$

where $D \triangleq -\mathbb{E}[\lambda^B] \cdot Imb$

The balancing payoff $z_B(\omega)$ can be expressed as follows:

- If $p > \lambda^B(\omega)$, then $z_B(\omega) = 0$
- If $p = \lambda^B(\omega)$, then $z_B(\omega) = (\lambda^B(\omega) - C) \cdot qa$ for some qa which is selected by the auctioneer. We get rid of this case by assuming that the auctioneer always activates zero MW of the supplier when the bid is at the money. Since this is a fringe supplier, the auctioneer can always source the imbalance energy from alternative suppliers. Thus, we have $qa(\omega) = 0$ and $z_B(\omega) = 0$ in this case.
- If $p < \lambda^B(\omega)$, then $z_B(\omega) = (\lambda^B(\omega) - C) \cdot q$

Therefore, the expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x>p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

where μ is the probability measure of the balancing price λ^B .

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B] - C) \cdot P^+ + D \\ C_2 &= \mathbb{E}[\lambda^B] - C \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \end{aligned}$$

In order to determine the optimal bidding strategy, let us first fix the bid quantity q of the agent. We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\ &= -(\mathbb{E}[\lambda^B] - C) + C_3(C) \\ &= -\left(\int_{x \leq C} (x - C) \cdot d\mu(x) + \int_{x > C} (x - C) \cdot d\mu(x)\right) + \int_{x > C} (x - C) \cdot d\mu(x) \\ &= -\int_{x \leq C} (x - C) \cdot d\mu(x) \\ &> 0 \end{aligned}$$

Therefore, it is optimal to bid $q^* = P^+$ in the balancing auction and $ai^* = 0$. This reflects the fact that, when being in active imbalance, the agent takes the risk of producing power when being out of the money. Instead, the balancing market will only activate the agent when its marginal cost is lower than the balancing price. The fact that the balancing and imbalance price are equal sends the correct incentive to the agent for bidding its entire capacity to the balancing auction. Note that every MW cleared in the day-ahead balancing capacity market comes with an obligation to bid that MW in the balancing energy auction, so this is a loss of opportunity because the agent does not have the chance to resort to active imbalance. For D1, since the optimal strategy of the agent is to bid its entire capacity in the balancing energy auction, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D1) Case 2: $\mathbb{E}[\lambda^B] < C, P^+ > 0, P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ > 0$ (and therefore $q \geq 0$):

$$\begin{aligned} \max_{ai} & (\mathbb{E}[\lambda^B] - C) \cdot ai - \mathbb{E}[\lambda^B] \cdot Imb \\ ai + q & \leq P^+ \\ ai & \geq 0 \end{aligned}$$

Since $\mathbb{E}[\lambda^B] < C$, we have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = D$$

where $D \triangleq -\mathbb{E}[\lambda^B] \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x>p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= D \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \end{aligned}$$

In order to determine the optimal bidding strategy, let us first fix the bid quantity q of the agent. We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost.

And, given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= C_3(C) \\ &= \int_{x>C} (x - C) \cdot d\mu(x) \\ &> 0 \end{aligned}$$

Therefore, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy auction, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D1) Case 3: $C \leq \mathbb{E}[\lambda^B]$, $P^+ = 0$, $P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} &\max_{ai} (C - \mathbb{E}[\lambda^B]) \cdot ai - \mathbb{E}[\lambda^B] \cdot Imb \\ &ai - q \leq -P^- \\ &ai \geq 0 \end{aligned}$$

Since $\mathbb{E}[\lambda^B] - C \geq 0$, we have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = D$$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x<p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= D \\ C_3(p) &= \int_{x<p} (x - C) \cdot d\mu(x) \end{aligned}$$

In order to determine the optimal bidding strategy, let us first fix the bid quantity q of the agent. We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= C_3(C) \\ &= \int_{x < C} (x - C) \cdot d\mu(x) \\ &< 0 \end{aligned}$$

Therefore, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D1) Case 4: $C > \mathbb{E}[\lambda^B], P^+ = 0, P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} &\max_{ai} (C - \mathbb{E}[\lambda^B]) \cdot ai - \mathbb{E}[\lambda^B] \cdot Imb \\ &ai - q \leq -P^- \\ &ai \geq 0 \end{aligned}$$

Since $\mathbb{E}[\lambda^B] - C \leq 0$, we have $ai^* = -P^- + q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B] - C) \cdot (P^- - q) + D$$

where $D \triangleq -\mathbb{E}[\lambda^B] \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x < p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B] - C) \cdot P^- + D \\ C_2 &= \mathbb{E}[\lambda^B] - C \\ C_3(p) &= \int_{x < p} (x - C) \cdot d\mu(x) \end{aligned}$$

In order to determine the optimal bidding strategy, let us first fix the bid quantity q of the agent. We can express the first-order conditions with respect to p as:

$$\begin{aligned}\frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q\end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned}\frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\ &= -(\mathbb{E}[\lambda^B] - C) + \int_{x < C} (x - C) \cdot d\mu(x) \\ &= -\left(\int_{x < C} (x - C) \cdot d\mu(x) + \int_{x \geq C} (x - C) \cdot d\mu(x)\right) \\ &\quad + \int_{x < C} (x - C) \cdot d\mu(x) \\ &= -\int_{x \geq C} (x - C) \cdot d\mu(x) \\ &< 0\end{aligned}$$

Therefore, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D1) Conclusion: We can state that it is always optimal for agents to bid their entire balancing capacity at the true marginal cost to the balancing energy market. For agents with upward balancing capacity ($P^+ > 0$), the opportunity cost of bidding their capacity to the day-ahead balancing capacity market is zero. We have characterized a pure strategy Nash equilibrium.

4.4.3 Design 2

In this section, we present the main results for (D2). The proofs for this design are available in appendix 4.7.1.

(D2) Case 1: $C \leq \mathbb{E}[\lambda^B]$, $P^+ > 0$, $P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D2) Case 2: $\mathbb{E}[\lambda^B] < C$, $P^+ > 0$, $P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D2) Case 3: $C \leq \mathbb{E}[\lambda^B], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D2) Case 4: $C > \mathbb{E}[\lambda^B], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D2) Conclusion: Under the assumption of independent symmetric imbalances, we can state that it is always optimal for agents to bid their entire balancing capacity at the true marginal cost to the balancing energy market. For agents with upward balancing capacity ($P^+ > 0$), the opportunity cost of bidding their capacity to the day-ahead balancing capacity market is zero. We have characterized a pure strategy Nash equilibrium.

4.4.4 Design 3

In this section, we present the main results for (D3). The proofs for this design are available in appendix 4.7.2.

(D3) Case 1: $C \leq \mathbb{E}[\lambda^B + \lambda^R] - \int_{x>C} (x - C) \cdot d\mu(x), P^+ > 0, P^- = 0$

In this case, it is optimal to bid $q^* = 0$ in the balancing energy market. This implies that for agents with low marginal costs, the incentive is to self-balance. The opportunity cost of bidding in the day-ahead balancing capacity market is equal to:

$$\begin{aligned} & \mathbb{E}[\lambda^B + \lambda^R] - C - \int_{x>C} (x - C) \cdot d\mu(x) \\ &= \mathbb{E}[\lambda^R] + \int (x - C) \cdot d\mu(x) - \int_{x>C} (x - C) \cdot d\mu(x) \\ &= \mathbb{E}[\lambda^R] + \int_{x \leq C} (x - C) \cdot d\mu(x) \\ &\leq \mathbb{E}[\lambda^R] \end{aligned}$$

(D3) Case 2: $\mathbb{E}[\lambda^B + \lambda^R] - \int_{x>C} (x - C) \cdot d\mu(x) < C \leq \mathbb{E}[\lambda^B + \lambda^R], P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D3) Case 3: $\mathbb{E}[\lambda^B + \lambda^R] < C, P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing auction. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D3) Case 4: $C \leq \mathbb{E}[\lambda^B + \lambda^R], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D3) Case 5: $C > \mathbb{E}[\lambda^B + \lambda^R], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D3) Conclusion: We can state that it is sometimes, but not always, optimal for agents to bid their entire balancing capacity at the true marginal cost to the balancing energy market. For agents with upward balancing capacity ($P^+ > 0$), the opportunity cost of bidding their capacity to the day-ahead balancing capacity market is less than or equal to the scarcity value $\mathbb{E}[\lambda^R]$. We have not characterized a pure strategy Nash equilibrium, since some agents find it optimal to self-balance. This design is depressing the scarcity price in two ways: (i) agents who find it optimal to bid their entire capacity to the balancing energy market face an opportunity cost of zero for bidding in the day-ahead balancing capacity market; and (ii) agents who find it optimal to self-balance face an opportunity cost which is less than the scarcity price $\mathbb{E}[\lambda^R]$. Moreover, the fact that some market participants find it optimal to perform self-balancing rather than offering their flexibility to the system might create a distortion of the price signal. Indeed, with certain agents performing self-balancing, it becomes difficult for a TSO to decide how much reserve to activate. This is due to the fact that the decision of which reserve to activate has to be reached before knowing how much self-balancing will take place. Therefore, there is no guarantee that the obtained dispatch will be optimal (respect the merit order) which would be the case if all the capacity was offered to the TSO as balancing energy. A second potential distortion of the price signal might come from the fact that, even if the TSO activates the right amount of reserve, if the most expensive activated unit is performing self-balancing, it is not part of the merit order and therefore does not set the price. This depresses the balancing price.

4.4.5 Design 4

In this section, we present the main results for (D4). The proofs for this design are available in appendix 4.7.3.

(D4) Case 1: $C \leq \mathbb{E}[\lambda^B], P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market.

(D4) Case 2: $\mathbb{E}[\lambda^B] < C, P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market.

(D4) Case 3: $C \leq \mathbb{E}[\lambda^B + \lambda^R], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D4) Case 4: $C \geq \mathbb{E}[\lambda^B + \lambda^R], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D4) Conclusion: We can state that it is always optimal for agents to bid their entire balancing capacity at the true marginal cost to the balancing energy market. We have characterized a pure strategy Nash equilibrium. The opportunity cost is $\mathbb{E}[\lambda^R]$. The opportunity cost can be explained by the real-time balancing capacity market. Indeed, the day-ahead contracted balancing capacity is bought back at a real-time price.

4.4.6 Design 5

In this section, we present the main results for (D5). The proofs for this design are available in appendix 4.7.4¹¹.

(D5) Case 1: $C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] - \int_{x>C} (x - C) \cdot d\mu(x)$, $P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = 0$ in the balancing energy market. This implies that for agents with low marginal costs, the incentive is to self-balance. The opportunity cost of bidding in the day-ahead balancing energy market is equal to $C_2 - C_3(C)$, which can be rewritten as:

$$\begin{aligned}
 & \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] - C - \int_{x>C} (x - C) \cdot d\mu(x) \quad (4.3) \\
 &= \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] + \int (x - C) \cdot d\mu(x) - \int_{x>C} (x - C) \cdot d\mu(x) \\
 &\leq \mathbb{E}[\lambda^R] + \mathbb{E}[\alpha^+ | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] + \int (x - C) \cdot d\mu(x) - \int_{x>C} (x - C) \cdot d\mu(x) \\
 &= \mathbb{E}[\lambda^R] + \int (x - C) \cdot d\mu(x) - \int_{x>C} (x - C) \cdot d\mu(x) \\
 &\leq \mathbb{E}[\lambda^R]
 \end{aligned}$$

We can compare this result with the one obtained for (D3). For (D3), we had the following opportunity cost:

$$\mathbb{E}[\lambda^B] + \mathbb{E}[\lambda^R] - C - \int_{x>C} (x - C) \cdot d\mu(x)$$

If we take the difference between the opportunity cost of (D3) and (D5), we obtain:

$$\mathbb{E}[\lambda^R] + \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] - \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s]$$

As shown in appendix 4.7.1, $\mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] = \mathbb{E}[\alpha^+ | \text{Imb}_{t-1}^s]$. Therefore, we can rewrite the difference as:

$$\mathbb{E}[\lambda^R] + \mathbb{E}[\alpha^+ | \text{Imb}_{t-1}^s] - \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s]$$

We are comparing the sum of two random variables with the maximum of these two random variables. As these two random variables are always positive, their sum is higher than their maximum.

This brings us to the conclusion that the opportunity cost given by (D5) is always smaller than or equal to the one given by (D3). Moreover, as $\mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s]$ do not cancel each other in Eq. (4.3), the profit is dependent on the previous period imbalance Imb_{t-1}^s .

(D5) Case 2: $\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] - \int_{x>C} (x - C) \cdot d\mu(x) < C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s]$, $P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market,

¹¹The notations of $\mathbb{E}[\alpha^+ | \text{Imb}_{t-1}^s]$ and $\mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s]$ are defined in appendix 4.7.1.

there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D5) Case 3: $C > \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s], P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D5) Case 4: $C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D5) Case 5: $C > \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D5) Conclusion: We can state that it is sometimes, but not always, optimal for agents to bid their entire balancing capacity at the true marginal cost to the balancing energy market. For agents with upward balancing capacity (P^+), the opportunity cost of bidding their capacity to the day-ahead balancing capacity market is less than or equal to the one of design (D3) which is itself lower than or equal to the scarcity value $\mathbb{E}[\lambda^R]$. We have not characterized a pure strategy Nash equilibrium, since some agents find it optimal to self-balance. This design is depressing the scarcity price in two ways: (i) agents who find it optimal to bid their entire capacity to the balancing energy market face an opportunity cost of zero for bidding in the day-ahead balancing capacity market; and (ii) agents who find it optimal to self-balance face an opportunity cost which is less than or equal to the opportunity cost in (D3) which is itself lower than the scarcity price $\mathbb{E}[\lambda^R]$. Moreover, the fact that some market participants find it optimal to perform self-balancing rather than offering their flexibility to the system might create a distortion of the price signal. Indeed, with certain agents performing self-balancing, it becomes difficult for a TSO to decide how much reserve to activate. This is due to the fact that the decision of which reserve to activate has to be taken before knowing how much self-balancing will take place. Therefore, there is no guarantee that the obtained dispatch will be optimal (respect the merit order) which would be the case if all the capacity was offered to the TSO as balancing energy. A second potential distortion of the price signal might come from the fact that, even if the TSO activates the right amount of reserve, if the most expensive activated unit is performing self-balancing, it is not part of the merit order and therefore does not set the price. This depresses the balancing price.

4.4.7 Design 6

In this section, we present the main results for (D6). The proofs for this design are available in appendix 4.7.5.

(D6) Case 1: $C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - \mathbb{E}[\lambda^R], P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market.

(D6) Case 2: $\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - \mathbb{E}[\lambda^R] < C, P^+ > 0, P^- = 0$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market.

(D6) Case 3: $C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D6) Case 4: $C > \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s], P^+ = 0, P^- < 0$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D6) Conclusion: We can state that it is always optimal for agents to bid their entire balancing capacity at the true marginal cost to the balancing energy market. The opportunity cost is $\mathbb{E}[\lambda^R]$. We have characterized a pure strategy Nash equilibrium. This result is equivalent to the one of design (D4). This is justified by the fact that (i) the only difference between the two designs is the pay-off from active imbalance. (ii) the optimal decision in both designs is to not perform active imbalance at all.

4.5 Case study

We now proceed to a numerical illustration in a simple case study. In section 4.5.1 we validate the analytical results of section 4.4 by considering a single fringe agent. In section 4.5.2 we assess the ability of the different designs to back-propagate balancing capacity prices by considering multiple agents that compete against each other.

4.5.1 Validation of Analytical Results

Consider a system with a fringe supplier that manages a flexible upward capacity of $P^+ = 1$ MW (and downward capacity of $P^- = 0$ MW). The marginal cost of the agent is $C = 50$ €/MWh. We discretize the action space as follows: the balancing energy auction bid q and balancing capacity auction bid q^R is either 0 MW or 1 MW, and the agent can bid any value p between 25 to 75 €/MWh, in increments of 5 €/MWh.

The system imbalance is assumed to be normally distributed with a mean of 0 MW and a standard deviation of 91.5 MW. The imbalance of the fringe agent is assumed to be uniformly distributed between $-0.5, 0$ and 0.5 . In the analytical model, the balancing supply function of the system is assumed to be affine, and is expressed mathematically as $a + b \cdot q$, where q is the amount of activated balancing capacity (with $q > 0$ corresponding to upward activation and $q < 0$ corresponding to downward activation), $a = 50$ €/MWh, and $b = 0.11$ (€/MWh)/MW. This supply function is an approximation of a balancing market with 8 agents, whose parameters are defined in Table 4.1. The fringe agent that we are interested in is agent A5.

	A1	A2	A3	A4	A5	A6	A7	A8
P^+	0	0	0	0	1	100	100	100
P^-	-100	-100	-100	-50	0	0	0	0
C	20	30	40	50	50	60	70	80

Table 4.1: The balancing capacity and marginal cost of different agents for the MDP code of section 4.5.1. Units are in [MW] for P^+ and P^- , and in [€/MWh] for C .

For the case of design (D2), we use the formula proposed by ELIA [ELI19]: $UI = LI = 150$ MW, and

$$\alpha^U = \alpha^D = \frac{200}{1 + \exp\left(\frac{450-x}{65}\right)} \quad (4.4)$$

where $x = \frac{|Imb^t| + |Imb_{t-1}^t|}{2}$ is the average of the absolute total system imbalances of the previous and current imbalance interval. For the case of design (D3) and (D4), we assume a value of $VOLL = 1000$ €/MWh.

Design	(D1)	(D3)	(D4)
q^* [MW]	1	0	1
p^* [€/MWh]	55	any	50
Average Profit [€]	6.34	14.43	18.85
Opportunity cost dR^*/dq [€/MWh]	0	8.11	12.71

Table 4.2: Results for (D1), (D3) and (D4) in the single-agent simulation.

Imb_{t-1}^t [MWh]	$(-\infty, -150]$	$(-150, 0]$	$(0, 150]$	$(150, \infty)$
q^* [MW]	1	1	1	1
p^* [€/MWh]	50	55	55	50
Average Profit [€]	6.43	6.30	6.32	6.46
dR^*/dq [€/MWh]	0	0	0	0

Table 4.3: Results for (D2) for different ranges of Imb_{t-1}^t in the single-agent simulation.

Imb_{t-1}^t [MWh]	$] \infty, -150]$	$[-150, 0]$	$[0, 150]$	$[150, \infty[$
q^* [MW]	0	0	0	0
p^* [€/MWh]	any	any	any	any
Average Profit [€]	13.56	14.03	14.02	14
dR^*/dq [€/MWh]	6.93	7.80	7.79	7.45

Table 4.4: Results for (D5) for different ranges of Imb_{t-1}^t in the single-agent simulation.

Imb_{t-1}^t [MWh]	$] \infty, -150]$	$[-150, 0]$	$[0, 150]$	$[150, \infty[$
q^* [MW]	1	1	1	1
p^* [€/MWh]	50	45	50	45
Average Profit [€]	18.98	18.58	18.90	19.15
dR^*/dq [€/MWh]	11.81	11.86	12.51	13.30

Table 4.5: Results for (D6) for different ranges of Imb_{t-1}^t in the single-agent simulation.

Design	(D1)	(D2)	(D3)	(D4)	(D6)
q^* [MW]	1	1	0	1	1
p^* [€/MWh]	50	50	any	50	50
Average Profit [€]	4.04	4.04	12.57	16.63	16.63
Opportunity cost dR^*/dq [€]	0	0	8.53	12.59	12.59

Table 4.6: Results for different market designs using the analytical solution.

For the single-agent simulation, we use the Q-learning algorithm¹² [WD92] under a uniformly distributed policy for the purpose of learning the Q function. We use a learning rate of $\frac{1}{n(s,a)}$ for each state-action pair (s, a) , where $n(s, a)$ counts the number of visits to (s, a) . We run 2,000,000 episodes for each design with the same seeds, in order to isolate

¹²This is a type of tabular method presented in section 1.4.2

Imb_{t-1}^t [MWh]	$] \infty, -150]$	$[-150, 0]$	$[0, 150]$	$[150, \infty[$
q^* [MW]	0	0	0	0
p^* [€/MWh]	any	any	any	any
Average Profit [€]	12.30	12.47	12.47	12.30
dR^*/dq [€/MWh]	8.24	8.41	8.41	8.24

Table 4.7: Results for (D5) for different ranges of Imb_{t-1}^t in the analytical solution.

the effect of the market design changes on the results.

We summarize the results of the simulations in Tables 4.2, 4.3, 4.4 and 4.5 as well as the analytical solution in Table 4.6 and 4.7. We observe the following. (i) For every design, the bid quantity and price are equivalent for the analytical case and the MDP model¹³. (ii) The profits are in the same range for the analytical solution and the MDP model. Differences (which amount to a range of 2 €) can be expected, because the analytical model assumes a continuous supply function, which is a continuous approximation of the stepwise supply function that is used in the MDP code. (iii) The opportunity costs are very close to each other for the analytical model and the MDP code. (iv) For designs (D2) and (D6), the value of the previous period imbalance, Imb_{t-1}^t , influences neither the selected action nor the profit, see Tables 4.3 and 4.5 4.6. These observations are in line with sections 4.4.3 and 4.4.7. For design (D5), the previous period, Imb_{t-1}^t , does not influence the selected action but it influences the profit and the opportunity cost as shown in Tables 4.4 and 4.7. This observation is in line with section 4.4.6.

4.5.2 Back-Propagation

We now concentrate on assessing experimentally the ability of the different market designs to back-propagate the real-time value of balancing capacity to the day-ahead balancing capacity market. For this purpose, we use our MDP model for developing a multi-agent simulation. In order to focus the analysis on the effects of the design in conditions of high competition for upward balancing capacity, we replace producers 5 – 8 by 35 producers with a capacity of 10 MW and marginal costs that increase uniformly from 50 €/MWh to 84 €/MWh.

We discretize the agent action space by having agents bid in price increments of 5 €/MWh and in quantity increments of half of their capacity. Each agent is facing a portfolio imbalance which is uniformly distributed between zero, half of its maximum capacity and minus half of its maximum capacity. There is also a system imbalance with a zero mean and a standard deviation of 72.9 MW. Agent imbalances are independent of each other and of the system imbalance. The day-ahead balancing capacity demand curve is assumed to be identical to the real-time balancing capacity demand curve, and based on the ORDC formula of Eq. (1.1).

¹³Indeed, in the MDP model, bidding at a price of 45 or 55 €/MWh is equivalent to bidding at 50 €/MWh, because there is no other producer with a marginal cost in the intervals $[45, 50)$ and $(50, 55]$. For design (D3) and (D5), the bid price does not matter, because the bid quantity is 0 MW.

We let every agent optimize its own policy using the Q-learning algorithm¹⁴ under an ϵ -greedy policy. During the learning phase, ϵ_k evolves as $\frac{0.05}{N-k}$, where N is the maximum number of iterations and k is the current iteration. Since all agents are learning simultaneously, from the perspective of any single agent, the environment is non-stationary, which implies that we have no convergence guarantees. In order to cope with the non-stationarity of the environment, we use a constant learning rate [SB18]. We run 1,500,000 iterations in blocks of 100. After each block of 100 iterations, we compute the outcome that we would have obtained in the balancing capacity market if each agent were applying its policy greedily. We plot the sample average of this balancing capacity price for the different designs in Fig. 4.2.

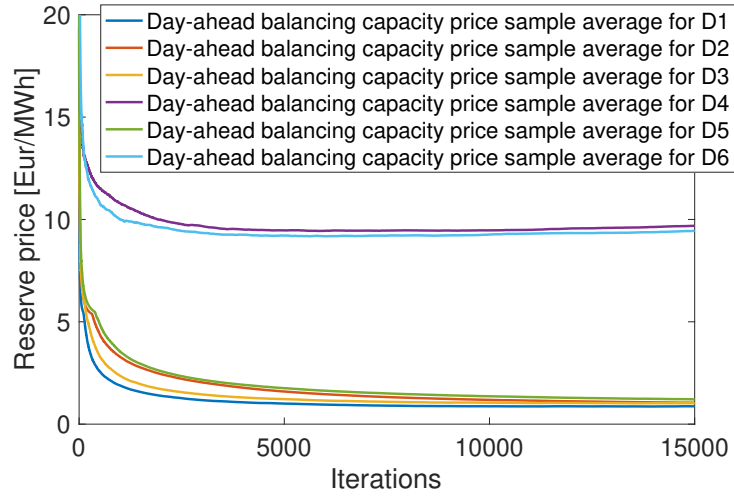


Figure 4.2: The evolution of the balancing capacity price in the simulation of section 4.5.2.

We observe the following. (i) For $(D1)$ and $(D2)$, the balancing capacity price sample average converges to a small value. This is anticipated by the analytical results, because the opportunity cost for each agent is equal to 0. (ii) For $(D3)$ and $(D5)$, the balancing capacity price sample average arrives slightly above the one resulting from $(D1)$. As shown analytically in section 4.4, under $(D3)$ and $(D5)$ certain low-cost producers may face a positive opportunity cost when bidding into the day-ahead balancing capacity market. Nevertheless, the resulting balancing capacity price remains close to the one of $(D1)$, because few producers are sufficiently cheap to fulfill this condition. (iii) Under design $(D4)$ and $(D6)$, the day-ahead balancing capacity price converges to a value which is close to the average real-time scarcity adder, i.e. 9.35 €/MWh.

4.5.3 Relaxing the Perfect Competition Assumption

The analytical model of section 4.4 assumes perfect competition. This is not necessarily representative of balancing markets, where balancing capacity requirements are sometimes quite small and the market may be dominated by a limited number of suppliers.

¹⁴An alternative method has been developed recently in order to solve this class of problems [Ber20]. The intuition is to generalize the value iteration algorithm to a multi-agent setting. To this aim, the author applies one step of the value iteration algorithm to each agent sequentially (the transition function for a precise agent can be computed if the policy of all the other agents is fixed). The advantage of this method compared to our approach is that it converges faster because it does not rely on sampling the uncertainty. Nevertheless, our approach has a natural interpretation as an iterative learning process that agents would engage in if we let them trade in the market without knowing the strategy of their competitors.

The analytical results of section 4.4 are only valid in a setting of perfect competition. Concretely, this assumption is required in order to arrive to the observation that balancing capacity prices are depressed under designs (D1), (D2), (D3) and (D5). If we lift the perfect competition assumption, then we can still use the MDP model of section 4.2.2 in order to investigate possible outcomes in the market. However, in such a setting it is typically difficult to verify that the point at which the MDP model converges is an actual equilibrium, since the condition that the agent policy should be optimal given the strategy of the other agents need to hold for every agent, and every possible state at every stage of the MDP model. Due to the fact that the Q functions are estimated in the MDP model, this verification is necessarily probabilistic, and typically accompanied by very weak confidence guarantees, since certain points of the state-action space are not explored extensively. In lieu of an analytical model that can predict equilibrium outcomes in the case of perfect competition, the results of the MDP model should therefore be considered as being purely suggestive because these results are only experimental and not based on an analytical characterization of an equilibrium. Moreover, we do not have an experimental guarantee that we converge to decisions that correspond to an equilibrium. Bearing this limitation in mind, we proceed with an application of our MDP model where we consider 11 agents. We maintain the 4 first agents of Table 4.1. We replace producers 5 – 8 of Table 4.1 by 7 agents with a capacity of 50 MW and marginal cost that increases uniformly from 50 to 80 €/MWh. We discretize the agent action space by having agents bid in price increments of 5 €/MWh and in quantity increments of a tenth of their capacity. The imbalance from the rest of the system has a zero mean and a standard deviation of $54.6MW$. The other parameters are kept the same as in the case of perfect competition.

We present the results of our simulation in Fig. 4.3. We observe that the price is higher for all designs compared to the case of perfect competition (see Fig. 4.2). This suggests that market power can be applied under every market design, and that our MDP model can be used for capturing such effects.

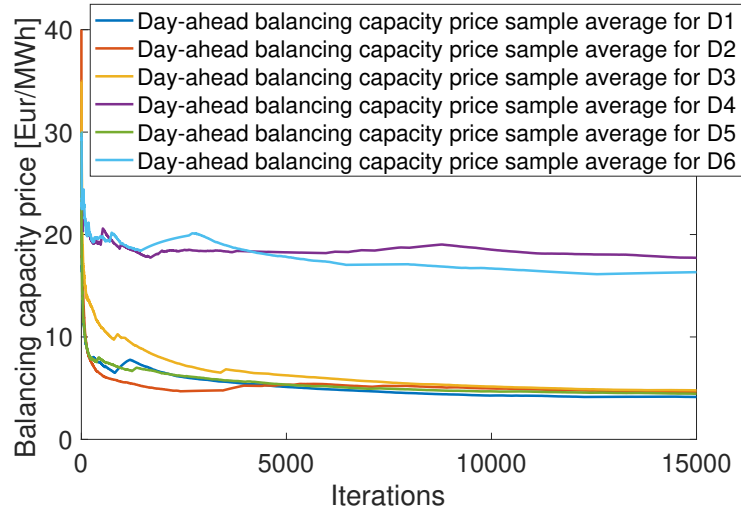


Figure 4.3: The evolution of the balancing capacity price in the simulation of section 4.5.3 under a setting of imperfect competition.

4.5.4 Other Factors Affecting Balancing Capacity Prices

The MDP model and analytical results that we have developed employ a number of simplifying assumptions. We discuss the assumption of perfect competition in section 4.5.3. In this section we comment on other factors that affect the formation of balancing capacity prices, including the inter-temporal coupling of market time units, fixed costs, and multiple balancing capacity types.

Inter-temporal coupling Both our analytical approach and our MDP model are implicitly assuming away inter-temporal dependencies. Inter-temporal dependencies occur in market clearing due to the dynamic constraints of resources (generator startups, ramp rates, min up / down times, storage levels, and so on) as well as the multi-interval nature of day-ahead and real-time energy and balancing capacity markets. For example, European day-ahead energy market clearing (as well as future integrated European day-ahead balancing capacity platforms, see articles 40-42 of [Eur17]) spans a 24-hour horizon. Similarly, a number of US day-ahead energy and balancing capacity markets based on co-optimization typically span a horizon of at least one day, while a number of US real-time markets such as CAISO and the New York ISO [Sch17] employ a multi-interval look ahead.

The introduction of inter-temporal coupling in our MDP model would create serious computational challenges that would require moving away from a simple lookup table representation of agent policies as explained in section 1.4.3. It is worth noting that pumped hydro resources in Belgium presently constitute a significant resource for the provision of frequency restoration reserves (a type of balancing capacity). The effect of inter-temporal constraints on Belgian market prices has been considered in [PS17, PSB18]. Inter-temporal constraints are ignored in this chapter in order to focus the analysis on the interaction of scarcity pricing and the back-propagation of balancing capacity prices to forward markets.

Fixed costs Fixed costs are not accounted for in our analysis. Belgium relies extensively on combined cycle gas turbines for frequency restoration reserves. These resources incur fixed costs for being online that contribute to the formation of forward balancing capacity prices [PS17]. The fixed cost associated to bringing a unit online so that it can provide balancing capacity to the system would introduce a non-zero cost associated to the sale of balancing capacity in the day-ahead market, and would therefore introduce a non-zero forward balancing capacity price that can contribute towards covering the operating cost that balancing capacity incurs for delivering balancing capacity services to the system. Instead, an important goal of scarcity pricing is to remunerate fixed long-run investment costs of resources that contribute to the system during scarcity. Our analysis uncovers balancing market designs that exhibit deficiencies in back-propagating this value to forward balancing capacity markets by considering the special case of zero fixed costs.

4.6 Conclusion

We present a methodology for analyzing the European balancing market based on an analytical derivation of optimal bidding under perfect competition assumptions, accompanied by an MDP-based simulation. The analysis exposes the inability of various market design alternatives in back-propagating the value of balancing capacity in day-ahead markets.

The analysis validates the ability of a real-time market for balancing capacity [PSd19] to back-propagate the value of balancing capacity to day-ahead markets, while also preserving the incentive of agents to make their balancing resources available in the balancing market.

In future research, we intend to analyse the results in presence of market power in more detail. Moreover, we are interested in further analyzing numerous important aspects of the mechanism. The specific parameter choices for computing the scarcity adders, i.e. the shape of the ORDC, are currently being investigated for the implementation of the mechanism in Belgium. Finally, it is important to understand the interaction of the mechanism with neighboring energy and balancing capacity markets that are not adopting the mechanism, and to ensure its compatibility with the legal framework of EBGL in this multi-area setting.

4.7 Appendix A: Computation of the Analytical Solution in the Perfect Competition Case

4.7.1 Design 2

Denote $Imb^t \triangleq Imb^s + Imb - ai^U + ai^D$ as the total system imbalance. The α penalty will be embedded in the imbalance price:

$$\begin{aligned}\lambda^I(\omega) &= \lambda^B(\omega) + \alpha(Imb^t) \cdot \mathbb{I}[Imb^t > UI] - \alpha(Imb^t) \cdot \mathbb{I}[Imb^t < LI] \\ &\simeq \lambda^B(\omega) + \alpha(Imb^s) \cdot \mathbb{I}[Imb^s > UI] - \alpha(Imb^s) \cdot \mathbb{I}[Imb^s < LI] \\ &= \lambda^B(\omega) + \alpha^+(\omega) - \alpha^-(\omega) \\ \alpha^+(\omega) &= \alpha(Imb^s) \cdot \mathbb{I}[Imb^s > UI] \\ \alpha^-(\omega) &= \alpha(Imb^s) \cdot \mathbb{I}[Imb^s < LI]\end{aligned}$$

Here Imb^s is the imbalance of the rest of the system (that does not include the agent), and $\alpha(x)$ is the α surcharge which applies when positive imbalances exceed the level UI , or when negative imbalances go below the level LI .

ELIA has decided to apply $UI = -LI = 150MW$, and the following formula for alpha:

$$\alpha(Imb^s | Imb_{t-1}^s) = \frac{200}{1 + \exp\left(\frac{450 - \frac{|Imb^s| + |Imb_{t-1}^s|}{2}}{65}\right)}$$

where we consider the imbalance of the previous balancing interval, Imb_{t-1}^s , as a fixed parameter. Note that the ELIA formula is symmetric, in the sense that $\alpha(-y|x) = \alpha(y|x)$ and $\alpha(y|x) = \alpha(y|-x)$. We also have $LI = -UI$. Therefore, the conditional expectation for the alpha penalties can be written as:

$$\begin{aligned}\mathbb{E}[\alpha^+ | Imb_{t-1}^s] &= \int_{y \leq LI} \alpha(y | Imb_{t-1}^s) \cdot \nu(y|x) \cdot dy \\ \mathbb{E}[\alpha^- | Imb_{t-1}^s] &= \int_{y \geq UI} \alpha(y | Imb_{t-1}^s) \cdot \nu(y|x) \cdot dy\end{aligned}$$

If we assume that (i) consecutive imbalances are independent and (ii) the imbalance distribution is symmetric. We obtain:

$$\begin{aligned}
\mathbb{E}[\alpha^- | Imb_{t-1}^s] &= \int_{y=-\infty}^{LI} \alpha(y | Imb_{t-1}^s) \cdot \nu(y) \cdot dy \\
&= - \int_{y'=-\infty}^{-LI} \alpha(-y' | Imb_{t-1}^s) \cdot \nu(-y') \cdot dy' \\
&= \int_{y'=-LI}^{\infty} \alpha(y' | Imb_{t-1}^s) \cdot \nu(y') \cdot dy' \\
&= \int_{y'=UI}^{\infty} \alpha(y' | Imb_{t-1}^s) \cdot \nu(y') \cdot dy' \\
&= \mathbb{E}[\alpha^+ | Imb_{t-1}^s]
\end{aligned}$$

(D2) Case 1: $C \leq \mathbb{E}[\lambda^B], P^+ > 0, P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ \geq 0$ (and therefore $q \geq 0$):

$$\begin{aligned}
&\max_{ai} (\mathbb{E}[\lambda^B] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - C) \cdot ai \\
&- (\mathbb{E}[\lambda^B] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot Imb \\
&ai + q \leq P^+ \\
&ai \geq 0
\end{aligned}$$

Under the assumption of independent symmetric imbalances, we have $\mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] = 0$, and the analysis reverts to that of (D1).

(D2) Case 2: $\mathbb{E}[\lambda^B] < C, P^+ > 0, P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ > 0$ (and therefore $q \geq 0$):

$$\begin{aligned}
&\max_{ai} (\mathbb{E}[\lambda^B] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - C) \cdot ai \\
&- (\mathbb{E}[\lambda^B] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot Imb \\
&ai + q \leq P^+ \\
&ai \geq 0
\end{aligned}$$

Under the assumption of independent symmetric imbalances, we have $\mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] = 0$, and the analysis reverts to that of (D1).

(D2) Case 3: $C \leq \mathbb{E}[\lambda^B], P^+ = 0, P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned}
& \max_{ai} (C - \mathbb{E}[\lambda^B] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot ai \\
& - (\mathbb{E}[\lambda^B] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot Imb \\
& ai - q \leq -P^- \\
& ai \geq 0
\end{aligned}$$

Under the assumption of independent symmetric imbalances, we have $\mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] = 0$, and the analysis reverts to that of (D1).

(D2) Case 4: $C > \mathbb{E}[\lambda^B], P^+ = 0, P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned}
& \max_{ai} (C - \mathbb{E}[\lambda^B] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot ai \\
& - (\mathbb{E}[\lambda^B] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot Imb \\
& ai - q \leq -P^- \\
& ai \geq 0
\end{aligned}$$

Under the assumption of independent symmetric imbalances, we have $\mathbb{E}[\alpha^+ | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] = 0$, and the analysis reverts to that of (D1).

4.7.2 Design 3

(D3) Case 1: $C \leq \mathbb{E}[\lambda^B + \lambda^R] - \int_{x>C} (x - C) \cdot d\mu(x), P^+ > 0, P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ \geq 0$ (and therefore $q \geq 0$):

$$\begin{aligned}
& \max_{ai} (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb \\
& ai + q \leq P^+ \\
& ai \geq 0
\end{aligned}$$

We have $ai^* = P^+ - q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot (P^+ - q) + E$$

where $E \triangleq -(\mathbb{E}[\lambda^B] + \mathbb{E}[\lambda^R]) \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned}
z_B &= \mathbb{E}[z_B(\omega)] \\
&= \int_{x>p} (x - C) \cdot q \cdot d\mu(x)
\end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned}
R(p, q) &= z_I + z_B \\
&= C_1 - C_2 \cdot q + C_3(p) \cdot q
\end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot P^+ + E \\ C_2 &= \mathbb{E}[\lambda^B + \lambda^R] - C \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\ &= -\mathbb{E}[\lambda^B + \lambda^R] + C + C_3(C) \\ &< 0 \end{aligned}$$

In this case, it is optimal to bid $q^* = 0$ in the balancing energy market. This implies that for agents with low marginal costs, the incentive is to self-balance. The opportunity cost of bidding in the day-ahead balancing capacity market is equal to:

$$\begin{aligned} &\mathbb{E}[\lambda^B + \lambda^R] - C - \int_{x>C} (x - C) \cdot d\mu(x) \\ &= \mathbb{E}[\lambda^R] + \int (x - C) \cdot d\mu(x) - \int_{x>C} (x - C) \cdot d\mu(x) \\ &= \mathbb{E}[\lambda^R] + \int_{x \leq C} (x - C) \cdot d\mu(x) \\ &\leq \mathbb{E}[\lambda^R] \end{aligned}$$

(D3) Case 2: $\mathbb{E}[\lambda^B + \lambda^R] - \int_{x>C} (x - C) \cdot d\mu(x) < C \leq \mathbb{E}[\lambda^B + \lambda^R]$, $P^+ > 0$, $P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ \geq 0$ (and therefore $q \geq 0$):

$$\begin{aligned} &\max_{ai} (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb \\ &ai + q \leq P^+ \\ &ai \geq 0 \end{aligned}$$

We have $ai^* = P^+ - q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot (P^+ - q) + E$$

where $E \triangleq -\mathbb{E}[\lambda^B + \lambda^R] \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x>p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot P^+ + E \\ C_2 &= \mathbb{E}[\lambda^B + \lambda^R] - C \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\ &= -\mathbb{E}[\lambda^B + \lambda^R] + C + C_3(C) \\ &= -\mathbb{E}[\lambda^B + \lambda^R] + C + \int_{x>C} (x - C) \cdot d\mu(x) \\ &> 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D3) Case 3: $\mathbb{E}[\lambda^B + \lambda^R] < C, P^+ > 0, P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ > 0$ (and therefore $q \geq 0$):

$$\begin{aligned}
& \max_{ai} (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb \\
& ai + q \leq P^+ \\
& ai \geq 0
\end{aligned}$$

We have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = E$$

where $E \triangleq -\mathbb{E}[\lambda^B + \lambda^R] \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned}
z_B &= \mathbb{E}[z_B(\omega)] \\
&= \int_{x>p} (x - C) \cdot q \cdot d\mu(x)
\end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned}
R(p, q) &= z_I + z_B \\
&= C_1 + C_3(p) \cdot q
\end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned}
C_1 &= E \\
C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x)
\end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned}
\frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\
&= -\mu(p) \cdot (p - C) \cdot q
\end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned}
\frac{\partial R(C, q)}{\partial q} &= C_3(C) \\
&= \int_{x>C} (x - C) \cdot d\mu(x) \\
&> 0
\end{aligned}$$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing auction. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D3) Case 4: $C \leq \mathbb{E}[\lambda^B + \lambda^R]$, $P^+ = 0$, $P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} & \max_{ai} (C - \mathbb{E}[\lambda^B + \lambda^R]) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb \\ & ai - q \leq -P^- \\ & ai \geq 0 \end{aligned}$$

Since $\mathbb{E}[\lambda^B + \lambda^R] - C \geq 0$, we have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = E$$

where $E \triangleq -(\mathbb{E}[\lambda^B + \lambda^R]) \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x < p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= E \\ C_3(p) &= \int_{x < p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= C_3(C) \\ &= \int_{x < C} (x - C) \cdot d\mu(x) \\ &< 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D3) Case 5: $C > \mathbb{E}[\lambda^B + \lambda^R]$, $P^+ = 0$, $P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} \max_{ai} & (C - \mathbb{E}[\lambda^B + \lambda^R]) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb \\ ai - q & \leq -P^- \\ ai & \geq 0 \end{aligned}$$

Since $\mathbb{E}[\lambda^B + \lambda^R] - C < 0$, we have $ai^* = -P^- + q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot (P^- - q) + E$$

where $E \triangleq -\mathbb{E}[\lambda^B + \lambda^R] \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x < p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot P^- + E \\ C_2 &= \mathbb{E}[\lambda^B + \lambda^R] - C \\ C_3(p) &= \int_{x < p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned}
\frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\
&= -(\mathbb{E}[\lambda^B + \lambda^R] - C) + \int_{x < C} (x - C) \cdot d\mu(x) \\
&= -\left(\int_{x < C} (x - C) \cdot d\mu(x) + \int_{x \geq C} (x - C) \cdot d\mu(x) \right) \\
&\quad + \int_{x < C} (x - C) \cdot d\mu(x) - \mathbb{E}[\lambda^R] \\
&= -\int_{x \geq C} (x - C) \cdot d\mu(x) - \mathbb{E}[\lambda^R] \\
&< 0
\end{aligned}$$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

4.7.3 Design 4

(D4) Case 1: $C \leq \mathbb{E}[\lambda^B]$, $P^+ > 0$, $P^- = 0$

The expected imbalance payoff (including the quantity of leftover capacity paid at the real-time price for balancing capacity) will be computed as follows for agents with $P^+ > 0$ (and therefore $q \geq 0$):

$$\begin{aligned}
&\max_{ai} (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb + \mathbb{E}[\lambda^R] \cdot (P^+ - ai - q) \\
&ai + q \leq P^+ \\
&ai \geq 0
\end{aligned}$$

This can be rewritten as:

$$\begin{aligned}
&\max_{ai} (\mathbb{E}[\lambda^B] - C) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb + (P^+ - q) \cdot \mathbb{E}[\lambda^R] \\
&ai + q \leq P^+ \\
&ai \geq 0
\end{aligned}$$

We have $ai^* = P^+ - q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B] - C + \mathbb{E}[\lambda^R]) \cdot (P^+ - q) + E$$

where $E \triangleq -\mathbb{E}[\lambda^B + \lambda^R] \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned}
z_B &= \mathbb{E}[z_B(\omega)] \\
&= \int_{x > p} (x - C) \cdot q \cdot d\mu(x) + \mathbb{E}[\lambda^R] \cdot q
\end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned}
R(p, q) &= z_I + z_B \\
&= C_1 - C_2 \cdot q + C_3(p) \cdot q + C_4 \cdot q
\end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot P^+ + E \\ C_2 &= \mathbb{E}[\lambda^B + \lambda^R] - C \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \\ C_4 &= \mathbb{E}[\lambda^R] \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q + C_4 \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) + C_4 \\ &= -\mathbb{E}[\lambda^B + \lambda^R] + C + C_3(C) + \mathbb{E}[\lambda^R] \\ &= -\mathbb{E}[\lambda^B] + C + C_3(C) \\ &= -\left(\int_{x \leq C} (x - C) \cdot d\mu(x) + \int_{x > C} (x - C) \cdot d\mu(x) \right) + \int_{x > C} (x - C) \cdot d\mu(x) \\ &= -\int_{x \leq C} (x - C) \cdot d\mu(x) \\ &\geq 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market.

(D4) Case 2: $\mathbb{E}[\lambda^B] < C, P^+ > 0, P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ > 0$ (and therefore $q \geq 0$):

$$\begin{aligned} &\max_{ai} (\mathbb{E}[\lambda^B] - C) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb + (P^+ - q) \cdot \mathbb{E}[\lambda^R] \\ &ai + q \leq P^+ \\ &ai \geq 0 \end{aligned}$$

We have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = E - F \cdot q$$

where $E \triangleq -\mathbb{E}[\lambda^B + \lambda^R] \cdot Imb + P^+ \cdot \mathbb{E}[\lambda^R]$ and $F = \mathbb{E}[\lambda^R]$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x>p} (x - C) \cdot q \cdot d\mu(x) + \mathbb{E}[\lambda^R] \cdot q \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q + C_4 \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= E \\ C_2 &= F \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \\ C_4 &= \mathbb{E}[\lambda^R] \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q + C_4 \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) + C_4 \\ &= -\mathbb{E}[\lambda^R] + C_3(C) + \mathbb{E}[\lambda^R] \\ &= C_3(C) \\ &= \int_{x>C} (x - C) \cdot d\mu(x) \\ &\geq 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market.

(D4) Case 3: $C \leq \mathbb{E}[\lambda^B + \lambda^R]$, $P^+ = 0$, $P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} &\max_{ai} (-\mathbb{E}[\lambda^B + \lambda^R] + C) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb \\ &ai - q \leq -P^- \\ &ai \geq 0 \end{aligned}$$

We have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = E$$

where $E \triangleq -\mathbb{E}[\lambda^B + \lambda^R] \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x < p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= E \\ C_3(p) &= \int_{x < p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= C_3(C) \\ &= \int_{x < C} (x - C) \cdot d\mu(x) \\ &\leq 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D4) Case 4: $C \geq \mathbb{E}[\lambda^B + \lambda^R]$, $P^+ = 0$, $P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} &\max_{ai} (-\mathbb{E}[\lambda^B + \lambda^R] + C) \cdot ai - \mathbb{E}[\lambda^B + \lambda^R] \cdot Imb \\ &ai - q \leq -P^- \\ &ai \geq 0 \end{aligned}$$

We have $ai^* = q - P^-$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot (P^- - q) + E$$

where $E \triangleq -\mathbb{E}[\lambda^B + \lambda^R] \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x < p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B + \lambda^R] - C) \cdot P^- + E \\ C_2 &= \mathbb{E}[\lambda^B + \lambda^R] - C \\ C_3(p) &= \int_{x < p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\ &= -(\mathbb{E}[\lambda^B + \lambda^R] - C) + \int_{x < C} (x - C) \cdot d\mu(x) \\ &= -\left(\int_{x < C} (x - C) \cdot d\mu(x) + \int_{x \geq C} (x - C) \cdot d\mu(x) \right) \\ &\quad + \int_{x < C} (x - C) \cdot d\mu(x) - \mathbb{E}[\lambda^R] \\ &= -\int_{x \geq C} (x - C) \cdot d\mu(x) - \mathbb{E}[\lambda^R] \\ &< 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

4.7.4 Design 5

(D5) Case 1: $C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - \int_{x>C} (x - C) \cdot d\mu(x)$, $P^+ > 0$, $P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ \geq 0$ (and therefore $q \geq 0$):

$$\begin{aligned} & \max_{ai} (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C) \cdot ai \\ & - (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot \text{Imb} \\ & ai + q \leq P^+ \\ & ai \geq 0 \end{aligned}$$

We have $ai^* = P^+ - q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C) \cdot (P^+ - q) + E$$

where $E \triangleq -(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot \text{Imb}$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x>p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C) \cdot P^+ + E \\ C_2 &= \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned}\frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\ &= -\mathbb{E}[\lambda^B] - \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] + \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] + C + C_3(C) \\ &< 0\end{aligned}$$

Therefore, it is optimal to bid $q^* = 0$ in the balancing auction. This implies that for agents with low marginal costs, the incentive is to self-balance. The opportunity cost of bidding in a balancing capacity auction is equal to $C_2 - C_3(C)$, which can be rewritten as:

$$\begin{aligned}&\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] - C - \int_{x>C} (x - C) \cdot d\mu(x) \\ &= \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] + \int (x - C) \cdot d\mu(x) - \int_{x>C} (x - C) \cdot d\mu(x) \\ &\leq \mathbb{E}[\lambda^R] + \mathbb{E}[\alpha^+ | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] + \int (x - C) \cdot d\mu(x) - \int_{x>C} (x - C) \cdot d\mu(x) \\ &= \mathbb{E}[\lambda^R] + \int (x - C) \cdot d\mu(x) - \int_{x>C} (x - C) \cdot d\mu(x) \\ &\leq \mathbb{E}[\lambda^R]\end{aligned}$$

(D5) Case 2: $\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] - \int_{x>C} (x - C) \cdot d\mu(x) < C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s], P^+ > 0, P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ \geq 0$ (and therefore $q \geq 0$):

$$\begin{aligned}&\max_{ai} (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] - C) \cdot ai \\ &- (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s]) \cdot \text{Imb} \\ &ai + q \leq P^+ \\ &ai \geq 0\end{aligned}$$

We have $ai^* = P^+ - q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] - C) \cdot (P^+ - q) + E$$

where $E \triangleq -(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s]) \cdot \text{Imb}$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned}z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x>p} (x - C) \cdot q \cdot d\mu(x)\end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned}R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q\end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C) \cdot P^+ + E \\ C_2 &= \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\ &= -\mathbb{E}[\lambda^B] - \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] + \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] + C + C_3(C) \\ &> 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D5) Case 3: $C > \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]$, $P^+ > 0$, $P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ \geq 0$ (and therefore $q \geq 0$):

$$\begin{aligned} &\max_{ai} (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C) \cdot ai \\ &\quad - (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot \text{Imb} \\ &ai + q \leq P^+ \\ &ai \geq 0 \end{aligned}$$

We have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = E$$

where $E \triangleq -(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot \text{Imb}$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x>p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= E \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= C_3(C) \\ &= \int_{x>C} (x - C) \cdot d\mu(x) \\ &> 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market. Since the optimal strategy of the agent is to bid its entire capacity in the balancing energy market, there is no opportunity cost for the agent. Therefore, the price at which the agent would bid in the day-ahead balancing capacity market is zero.

(D5) Case 4: $C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s], P^+ = 0, P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} &\max_{ai} (C - \mathbb{E}[\lambda^B] - \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] + \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot ai \\ &\quad - (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot \text{Imb} \\ &ai - q \leq -P^- \\ &ai \geq 0 \end{aligned}$$

Since $\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C \geq 0$, we have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = E$$

where $E \triangleq -(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot \text{Imb}$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x < p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= E \\ C_3(p) &= \int_{x < p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= C_3(C) \\ &= \int_{x < C} (x - C) \cdot d\mu(x) \\ &< 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D5) Case 5: $C > \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]$, $P^+ = 0$, $P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} &\max_{ai} (C - \mathbb{E}[\lambda^B] - \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] + \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot ai \\ &- (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot Imb \\ ai - q &\leq -P^- \\ ai &\geq 0 \end{aligned}$$

Since $\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - C \leq 0$, we have $ai^* = -P^- + q$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - C) \cdot (P^- - q) + E$$

where $E \triangleq -(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|Imb_{t-1}^s] - \mathbb{E}[\alpha^-|Imb_{t-1}^s]) \cdot Imb$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x < p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|Imb_{t-1}^s] - \mathbb{E}[\alpha^-|Imb_{t-1}^s] - C) \cdot P^- + E \\ C_2 &= (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|Imb_{t-1}^s] - \mathbb{E}[\alpha^-|Imb_{t-1}^s] - C) \\ C_3(p) &= \int_{x < p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned}
\frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\
&= -(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - C) \\
&\quad + \int_{x < C} (x - C) \cdot d\mu(x) \\
&= -\left(\int_{x < C} (x - C) \cdot d\mu(x) + \int_{x \geq C} (x - C) \cdot d\mu(x) \right) \\
&\quad + \int_{x < C} (x - C) \cdot d\mu(x) - \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] + \mathbb{E}[\alpha^- | Imb_{t-1}^s] \\
&= -\int_{x \geq C} (x - C) \cdot d\mu(x) - \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] + \mathbb{E}[\alpha^- | Imb_{t-1}^s] \\
&= -\int_{x \geq C} (x - C) \cdot d\mu(x) - \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] \\
&\leq -\int_{x \geq C} (x - C) \cdot d\mu(x) - \mathbb{E}[\alpha^+ | Imb_{t-1}^s] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] \\
&= -\int_{x \geq C} (x - C) \cdot d\mu(x) \\
&< 0
\end{aligned}$$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

4.7.5 Design 6

(D6) Case 1: $C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - \mathbb{E}[\lambda^R], P^+ > 0, P^- = 0$

The expected imbalance payoff (including the quantity of leftover capacity paid at the real-time price for balancing capacity) will be computed as follows for agents with $P^+ > 0$ (and therefore $q \geq 0$):

$$\begin{aligned}
&\max_{ai} (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - C) \cdot ai \\
&\quad - (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot Imb \\
&\quad + \mathbb{E}[\lambda^R] \cdot (P^+ - ai - q) \\
&ai + q \leq P^+ \\
&ai \geq 0
\end{aligned}$$

This can be rewritten as:

$$\begin{aligned}
&\max_{ai} (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - C - \mathbb{E}[\lambda^R]) \cdot ai \\
&\quad - (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s]) \cdot Imb + (P^+ - q) \cdot \mathbb{E}[\lambda^R] \\
&ai + q \leq P^+ \\
&ai \geq 0
\end{aligned}$$

We have $ai^* = P^+ - q$. The expected payoff z_I is then expressed as follows:

$$z_I = \left(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C \right. \\ \left. - \mathbb{E}[\lambda^R] + \mathbb{E}[\lambda^R] \right) \cdot (P^+ - q) + E$$

where $E \triangleq - \left(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] \right) \cdot \text{Imb}$

The expected balancing payoff can be expressed as follows:

$$z_B = \mathbb{E}[z_B(\omega)] \\ = \int_{x>p} (x - C) \cdot q \cdot d\mu(x) + \mathbb{E}[\lambda^R] \cdot q$$

The overall payoff of the agent can therefore be expressed as follows:

$$R(p, q) = z_I + z_B \\ = C_1 - C_2 \cdot q + C_3(p) \cdot q + C_4 \cdot q$$

where the terms can be described as follows:

$$C_1 = \left(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C \right) \cdot P^+ + E \\ C_2 = \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C \\ C_3(p) = \int_{x>p} (x - C) \cdot d\mu(x) \\ C_4 = \mathbb{E}[\lambda^R]$$

We can express the first-order conditions with respect to p as:

$$\frac{\partial R(p, q)}{\partial p} = C'_3(p) \cdot q \\ = -\mu(p) \cdot (p - C) \cdot q$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q + C_4 \cdot q$$

If we take the derivative with respect to q , we have:

$$\frac{\partial R(C, q)}{\partial q} = -C_2 + C_3(C) + C_4 \\ = -\mathbb{E}[\lambda^B] - \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] + \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] + C + C_3(C) + C_4 \\ = -\mathbb{E}[\lambda^B] - \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] + \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] + C + C_3(C) + \mathbb{E}[\lambda^R] \\ \geq -\mathbb{E}[\lambda^B] + C + C_3(C) \\ = - \left(\int_{x \leq C} (x - C) \cdot d\mu(x) + \int_{x > C} (x - C) \cdot d\mu(x) \right) + \int_{x > C} (x - C) \cdot d\mu(x) \\ \geq 0$$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market.

(D6) Case 2: $\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|Imb_{t-1}^s] - \mathbb{E}[\alpha^-|Imb_{t-1}^s] - \mathbb{E}[\lambda^R] < C, P^+ > 0, P^- = 0$

The expected imbalance payoff will be computed as follows for agents with $P^+ > 0$ (and therefore $q \geq 0$):

$$\begin{aligned} & \max_{ai} \left(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|Imb_{t-1}^s] - \mathbb{E}[\alpha^-|Imb_{t-1}^s] - C - \mathbb{E}[\lambda^R] \right) \cdot ai \\ & - \left(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|Imb_{t-1}^s] - \mathbb{E}[\alpha^-|Imb_{t-1}^s] \right) \cdot Imb + (P^+ - q) \cdot \mathbb{E}[\lambda^R] \\ & ai + q \leq P^+ \\ & ai \geq 0 \end{aligned}$$

We have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = E - F \cdot q$$

where

- $E \triangleq - \left(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|Imb_{t-1}^s] - \mathbb{E}[\alpha^-|Imb_{t-1}^s] \right) \cdot Imb + P^+ \cdot \mathbb{E}[\lambda^R]$
- $F = \mathbb{E}[\lambda^R]$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x>p} (x - C) \cdot q \cdot d\mu(x) + \mathbb{E}[\lambda^R] \cdot q \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q + C_4 \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= E \\ C_2 &= F \\ C_3(p) &= \int_{x>p} (x - C) \cdot d\mu(x) \\ C_4 &= \mathbb{E}[\lambda^R] \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= -\mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q + C_4 \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned}
\frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) + C_4 \\
&= -\mathbb{E}[\lambda^R] + C_3(C) + \mathbb{E}[\lambda^R] \\
&= C_3(C) \\
&= \int_{x>C} (x - C) \cdot d\mu(x) \\
&\geq 0
\end{aligned}$$

In this situation, it is optimal to bid $q^* = P^+$ in the balancing energy market.

(D6) Case 3: $C \leq \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s], P^+ = 0, P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned}
&\max_{ai} \left(-\mathbb{E}[\lambda^B] - \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] + \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] + C \right) \cdot ai \\
&- \left(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] \right) \cdot \text{Imb} \\
&ai - q \leq -P^- \\
&ai \geq 0
\end{aligned}$$

We have $ai^* = 0$. The expected payoff z_I is then expressed as follows:

$$z_I = E$$

where $E \triangleq - \left(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | \text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^- | \text{Imb}_{t-1}^s] \right) \cdot \text{Imb}$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned}
z_B &= \mathbb{E}[z_B(\omega)] \\
&= \int_{x<p} (x - C) \cdot q \cdot d\mu(x)
\end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned}
R(p, q) &= z_I + z_B \\
&= C_1 + C_3(p) \cdot q
\end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned}
C_1 &= E \\
C_3(p) &= \int_{x<p} (x - C) \cdot d\mu(x)
\end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned}
\frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\
&= \mu(p) \cdot (p - C) \cdot q
\end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at

C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= C_3(C) \\ &= \int_{x < C} (x - C) \cdot d\mu(x) \\ &\leq 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

(D6) Case 4: $C > \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]$, $P^+ = 0$, $P^- < 0$

The expected imbalance payoff will be computed as follows for agents with $P^- < 0$ (and therefore $q \leq 0$):

$$\begin{aligned} &\max_{ai} (-\mathbb{E}[\lambda^B] - \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] + \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] + C) \cdot ai \\ &- (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot \text{Imb} \\ &ai - q \leq -P^- \\ &ai \geq 0 \end{aligned}$$

We have $ai^* = q - P^-$. The expected payoff z_I is then expressed as follows:

$$z_I = (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C) \cdot (P^- - q) + E$$

where $E \triangleq -(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s]) \cdot \text{Imb}$

The expected balancing payoff can be expressed as follows:

$$\begin{aligned} z_B &= \mathbb{E}[z_B(\omega)] \\ &= \int_{x < p} (x - C) \cdot q \cdot d\mu(x) \end{aligned}$$

The overall payoff of the agent can therefore be expressed as follows:

$$\begin{aligned} R(p, q) &= z_I + z_B \\ &= C_1 - C_2 \cdot q + C_3(p) \cdot q \end{aligned}$$

where the terms can be described as follows:

$$\begin{aligned} C_1 &= (\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C) \cdot P^- + E \\ C_2 &= \mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+)|\text{Imb}_{t-1}^s] - \mathbb{E}[\alpha^-|\text{Imb}_{t-1}^s] - C \\ C_3(p) &= \int_{x < p} (x - C) \cdot d\mu(x) \end{aligned}$$

We can express the first-order conditions with respect to p as:

$$\begin{aligned} \frac{\partial R(p, q)}{\partial p} &= C'_3(p) \cdot q \\ &= \mu(p) \cdot (p - C) \cdot q \end{aligned}$$

We note that the payoff function $R(p, q)$ for fixed q is increasing in $(-\infty, C]$, zero at C , and decreasing in $[C, +\infty)$. Thus, for any q , an optimal strategy is to bid the true cost. And, given this strategy, the payoff becomes

$$R(C, q) = C_1 - C_2 \cdot q + C_3(C) \cdot q$$

If we take the derivative with respect to q , we have:

$$\begin{aligned} \frac{\partial R(C, q)}{\partial q} &= -C_2 + C_3(C) \\ &= -(\mathbb{E}[\lambda^B] + \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] - \mathbb{E}[\alpha^- | Imb_{t-1}^s] - C) \\ &\quad + \int_{x < C} (x - C) \cdot d\mu(x) \\ &= -\left(\int_{x < C} (x - C) \cdot d\mu(x) + \int_{x \geq C} (x - C) \cdot d\mu(x) \right) \\ &\quad + \int_{x < C} (x - C) \cdot d\mu(x) - \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] + \mathbb{E}[\alpha^- | Imb_{t-1}^s] \\ &= -\int_{x \geq C} (x - C) \cdot d\mu(x) - \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] + \mathbb{E}[\alpha^- | Imb_{t-1}^s] \\ &= -\int_{x \geq C} (x - C) \cdot d\mu(x) - \mathbb{E}[\max(\lambda^R, \alpha^+) | Imb_{t-1}^s] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] \\ &\leq -\int_{x \geq C} (x - C) \cdot d\mu(x) - \mathbb{E}[\alpha^+ | Imb_{t-1}^s] + \mathbb{E}[\alpha^+ | Imb_{t-1}^s] \\ &= -\int_{x \geq C} (x - C) \cdot d\mu(x) \\ &< 0 \end{aligned}$$

In this situation, it is optimal to bid $q^* = P^-$ in the balancing energy market.

Chapter 5

Conclusions and Perspectives

This thesis focuses on the analysis of ways for improving the remuneration of flexible assets in electricity markets. In chapter 2 we develop trading strategies for a storage unit in the CIM. We model this problem as an MDP. In order to make this problem tractable, we rely on policy function approximation. We parametrize our policy using relevant exogenous variables. We demonstrate that our approach outperforms a rolling intrinsic method which is commonly used in the industry. Then, in chapter 3, we compute an analytical solution for trading a fixed quantity of power in a simplified version of the CIM. We use this analytical solution in order to parametrize a value function approximation algorithm that can trade in more realistic settings. We compare the performance of our value function approximation approach and of an alternative SDDP approach on the simplified version of the CIM. Finally, in chapter 4, we analyse market design options for the introduction of a scarcity pricing mechanism in Belgium. We develop an analytical framework in order to analyse the behaviour of a fringe agent under different balancing market variants. In order to validate these analytical results, we use a multi-agent Reinforcement Learning simulation. In the following, we first present a summary of our conclusions and then discuss a list of areas for future work that have been inspired by the present research.

5.1 A Summary of Conclusions

5.1.1 Case Study on a storage unit trading in the CIM

- (a) **Performance of our threshold policy:** Our proposed approach outperforms the rolling intrinsic policy, which is commonly employed in practice for storage units, by increasing profitability by 17.8% on out-of-sample testing for a storage unit with perfect round-trip efficiency and by 13.6% for a storage unit with a round-trip efficiency of 81%. Moreover, our policy is consistently performing better than rolling intrinsic by generating higher profit on 77.4% of the days with perfect round-trip efficiency and 64.6% with a round-trip efficiency of 81%.
- (b) **Trading frequency:** The frequency at which we trade has a strong impact on the profit. When trading at a one-second frequency, our method performs 19.2% better than at a 1-hour frequency. We observe that, at high frequency, two factors contribute to the profit: (i) significant arbitrage possibilities of the storage unit that are predictable, and (ii) smaller arbitrages that cannot be anticipated. We demonstrate that our method is efficient in capturing both of these gains, whereas

rolling intrinsic is effective at capturing small arbitrages but inefficient at capturing the large predictable arbitrages. This is driven by the fact that rolling intrinsic sells its flexibility whenever a profitable trade arrives, in a short-sighted fashion.

5.1.2 Case Study on the trading of a fixed quantity in the CIM

- (a) **Computation of an analytical solution:** We propose a simplified model for the CIM. We compute analytically the optimal policy and the optimal value function for trading a fixed quantity of power in this simplified version of the CIM.
- (b) **Comparison of Value Function Approximation and Stochastic Dual Dynamic Programming:** We demonstrate that both our value function approximation and SDDP are able to recover the optimal value function and the optimal policy on a 10-step example. Therefore, both of these methods can be considered for implementation on the real CIM.

5.1.3 Case Study on the market design of a scarcity pricing mechanism in Belgium

- (a) **Real-time market for reserve:** Our analysis validates the ability of a real-time market for balancing capacity [PSd19] to back-propagate the value of balancing capacity to day-ahead markets, while also preserving the incentive of agents to make their balancing resources available in the balancing market.
- (b) **Alternatives without a real-time market for reserve:** Our analysis illustrates that the alternative approaches are not able to back-propagate adequately the value of balancing capacity to day-ahead markets. Moreover, we show that any obtained back-propagation is at the expense of having agents bidding in the balancing market at their marginal cost.
- (c) **Non-zero day-ahead balancing capacity prices in the current design:** Our results are not incompatible with the observation of non-zero reserve prices for balancing capacity in practice. Indeed, our conclusion is that there is a weak back-propagation of scarcity to the day-ahead balancing capacity market (and not that the day-ahead reserve price should always be equal to 0 €/MWh). One of the explanations for these non-zero prices relates to the technical constraints of the power plants that are not modelled in our example. Specifically, power plants have minimum production levels¹ and start-up costs. Another explanation could be the fact that financial penalties are applied if a BSP is not able to deliver the promised reserve. Therefore, a BSP may always ask for a compensation for the risk of being exposed to these penalties. A last explanation could be non-competitive behaviour due to the fact that there are few participants in the balancing capacity markets.

5.2 Future Areas of Research

- (a) **Value of coordination with earlier markets:** In chapter 2, we consider trading exclusively in the CIM. An important extension would be to compute the value

¹The minimum production constraint might force the agent to operate its power plant while the day-ahead price is below its marginal cost, in order to be able to provide reserve. In order to compensate this loss, the BSP can ask for a non-zero price in the balancing capacity market.

of coordinating trading strategies between early markets (day-ahead and intraday auction, day-ahead reserve markets) and the CIM. Day-ahead reserve markets seem to be especially appropriate candidates because they also correspond to a promising revenue stream for flexible assets such as a storage unit.

- (b) **Multi-reservoir problem:** In chapter 2, we consider only one reservoir. It would be interesting to extend our results to several connected reservoirs, as in [BJF14].
- (c) **Test our trading strategy for a wind unit in the real CIM:** In chapter 3, we only compare the performance of SDDP and VFA for trading a fixed quantity of power on a simplified version of the CIM. It is an important extension to assess the performance of both of these methods in the real CIM. This includes analysing in detail the relevant exogenous parameters that can influence the evolution of the CIM price and how to include such parameters in the value function parametrization. This step would allow us to realize a full-blown test case and draw definitive conclusions, as in chapter 2.
- (d) **Market power in scarcity pricing:** In chapter 4, we present the results of our multi-agent simulation, where we relax the assumption of perfect competition. It would be interesting to verify if our simulation converges to an equilibrium. Towards this aim, we could develop an analytical framework, as for the perfect competition case. An alternative approach would be to check, after convergence for each agent, whether the policy of each agent is optimal when holding the policy of the other agents fixed and equal to their values at convergence [YQLS19].
- (e) **Multiple reserve products:** In chapter 4, we consider the case of a single reserve product. In reality, there are three reserve products² in Belgium: (i) automatic frequency restoration reserve (a reserve that reacts automatically, following a controller, to frequency deviations on a 4-second basis), (ii) manual frequency restoration reserve for scheduled activation (reserve that is activated manually for a delivery period of 15 minutes and that can only be activated at the beginning of the interval), and (iii) manual frequency restoration reserve for direct activation (reserve that can be activated at any moment of the delivery interval and that should remain activated during its activation interval and the following one). The presence of multiple reserve products implies different scarcity adders. It is important to define precisely which scarcity adder applies to each reserve product, and to energy, in order to ensure correct incentives for BSPs.
- (f) **Approximation of the co-optimization of energy and reserve:** The formula of the scarcity adder, presented in chapter 4, is an approximation of a co-optimization of energy and reserve in real time. An interesting question would be to analyze in which situations the approximation is exact and in which situations the approximation produces a different result compared to an explicit co-optimization.
- (g) **Introduction of scarcity pricing in adequacy studies:** This is an important research topic because Belgium is currently implementing a capacity remuneration mechanism. The rationale for this implementation is based on adequacy studies that demonstrate that the profit of new power plants would be insufficient to incentivize

²We ignore frequency containment reserves, which are out of scope.

investments. Nevertheless, these adequacy studies do not account for the impact of a scarcity pricing mechanism on the profitability of power plants³. A potential way for including scarcity pricing in adequacy studies would be to consider the idea used in the simulator of [CP21]. Specifically, rather than solving one day-ahead unit commitment model, the TSO could solve several sequential optimization models (day-ahead, intraday, pre-real-time, real-time) in order to better approximate the real-time operations of the system.

- (h) **European balancing platforms:** European balancing markets are becoming integrated with the upcoming introduction of the MARI (for manual frequency restoration reserve) and Picasso (for automatic frequency restoration reserve) platforms. The idea is that the offers of the BSPs from different countries will compete in centralized European platforms, in order to reduce the cost of procurement. In this context, it is important to assess the impact, on neighbouring countries, of a unilateral implementation of scarcity pricing in Belgium. Specifically, we need to ensure that the introduction of a scarcity pricing mechanism in Belgium does not distort competition between Belgian BSPs and BSPs from neighbouring countries.

³Notice that this is expressly intended in articles 20(3) and 23(5) of regulation 2019/943 [Eur19].

Bibliography

- [ACE20] ACER. Evaluation of responses to the public consultation on the methodology to determine prices for the balancing energy that results from the activation of balancing energy bids, 2020. URL: https://www.acer.europa.eu/Official_documents/Acts_of_the_Agency/Annexes%20to%20the%20DECISION%20OF%20THE%20AGENCY%20FOR%20THE%20C2/ACER%20Decision%20on%20the%20Methodology%20for%20pricing%20balancing%20energy%20-%20Annex%20II.pdf.
- [Bal18] C. Balardy. An empirical analysis of the bid-ask spread in the German power continuous market, 2018. URL: http://www.ceem-dauphine.org/assets/wp/pdf/0918-CEEM_Working_Paper_35_Clara_BALARDY.pdf.
- [BB01] J. Bower and D. Bunn. Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the England and Wales electricity market. *Journal of Economic Dynamics and Control*, 25(3-4):561–592, Mar. 2001.
- [BEPC18] I. Boukas, D. Ernst, A. Papavasiliou, and B. Cornelusse. Intra-day bidding strategies for storage devices using deep reinforcement learning. In *15th International Conference on the European Energy Market (EEM)*, 2018.
- [Ber20] D. Bertsekas. Multiagent value iteration algorithms in dynamic programming and reinforcement learning. *Results in Control and Optimization*, 1:100003, 2020.
- [BET⁺19] I. Boukas, D. Ernst, T. Theate, A. Bolland, A. Huynen, M. Buchwald, C. Wynants, and B. Cornelusse. A deep reinforcement learning framework for continuous intraday market bidding, 2019. URL: <https://arxiv.org/pdf/2004.05940.pdf>.
- [BH16] S. Braun and R. Hoffmann. Intraday optimization of pumped hydro power plants in the German electricity market. *Energy Procedia*, 87:45–52, Jan. 2016.
- [BJF14] T. Boomsma, N. Juul, and S. Fleten. Bidding in sequential electricity markets: The Nordic case. *European Journal of Operational Research*, 238(3):797–809, Nov. 2014.
- [BLS⁺12] D. Bertsimas, E. Litvinov, A. Sun, J. Zhao, and T. Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, 2012.

- [BO01] D. Bunn and F. Oliveira. Agent-based simulation-an application to the new electricity trading arrangements of England and Wales. *Journal of Economic Dynamics and Control*, 8(3-4):493–503, Oct. 2001.
- [BP18] G. Bertrand and A. Papavasiliou. An analysis of threshold policies for trading in continuous intraday electricity markets. In *15th International Conference on the European Energy Market (EEM)*, 2018.
- [BP19a] G. Bertrand and A. Papavasiliou. A reinforcement-learning based threshold policies for continuous intraday electricity market trading. In *IEEE PES General Meeting*, 2019.
- [BP19b] G. Bertrand and A. Papavasiliou. Adaptive trading in continuous intraday electricity markets for a storage unit. *IEEE Transactions on Power Systems*, 35(3):2339–2350, Dec. 2019.
- [Bra16] S. Braun. Hydropower storage optimization considering spot and intraday auction market. *Energy Procedia*, 87:36–44, Jan. 2016.
- [CAI11] CAISO. Flexible ramping products. Technical report, CAISO, 2011. URL: <http://www.caiso.com/Documents/FlexibleRampingProductStrawProposal.pdf>.
- [CHR13] J. Chaves, R. Hakvoort, and A. Ramos. Short-term strategies for Dutch wind power producers to reduce imbalance costs. *Energy Policy*, 52:573–582, Jan. 2013.
- [CJP15] A. Cartea, S. Jaimungal, and J. Penalva. *Algorithmic and high-frequency trading*. Cambridge University Press, 2015.
- [CP21] J. Cartuyvels and A. Papavasiliou. A market simulation methodology for the calibration of ORDC, 2021.
- [Cra17] P. Cramton. Electricity market design. *Oxford Review of Economic Policy*, 33(4):589–612, 2017.
- [CREa] CREG. Note on scarcity pricing applied to Belgium. URL: <http://www.creg.info/pdf/Divers/Z1527EN.pdf>.
- [CREb] CREG. Note relative aux évolutions marquantes sur les marchés de gros belges de l’électricité et du gaz naturel en 2020. URL: <https://www.creg.be/sites/default/files/assets/Publications/Notes/Z2187FR.pdf>.
- [CREc] CREG. Study on the functioning and price evolution of the Belgian wholesale electricity market monitoring report 2018. URL: <https://www.creg.be/sites/default/files/assets/Publications/Studies/F1958EN.pdf>.
- [DDB20] A. Downward, O. Dowson, and R. Baucke. Stochastic dual dynamic programming with stagewise dependent objective uncertainty. *Operations Research Letters*, 48(1):33–39, Jan. 2020.

- [DJS99] S. Deng, B. Johnson, and A. Sogomonian. Spark spread options and the valuation of electricity generation assets. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999.
- [DK20] O. Dowson and L. Kapelevich. SDDP.jl: a Julia package for stochastic dual dynamic programming. *INFORMS Journal on Computing*, 2020.
- [DO03] S. Deng and S. Oren. Incorporating operational characteristics and startup costs in option-based valuation of power generation capacity. *Probability in the Engineering and Informational Sciences*, 17(2):155–181, 2003.
- [EC] European Commission. 2020 climate and energy package. URL: https://ec.europa.eu/clima/policies/strategies/2020_en.
- [EC08] European Commission. Memo on the renewable energy and climate change package, Jan. 2008.
- [ELIa] ELIA. Auction calendar. URL: <https://www.elia.be/en/grid-data/balancing/capacity-auction-calendar>.
- [ELIb] ELIA. Belgian electricity market: Implementation plan. URL: <https://ec.europa.eu/energy/sites/ener/files/belgian-electricity-market-implementation-plan.pdf>.
- [ELIc] ELIA. Mfrr 2020 product design note. URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjMr7e0k63wAhWFH-wKHVJ4DpUQFjAAegQIAxAD&url=https%3A%2F%2Fwww.elia.be%2F-%2Fmedia%2Fproject%2Felia%2Felia-site%2Felectricity-market-and-system---document-library%2Fbalancing---balancing-services-and-bsp%2F2019%2F2019-design-note-mfrr-2020_version-01032019.pdf&usg=A0vVaw3xBQ_jIskuZDsY_GU6SA83.
- [ELId] ELIA. Mfrr 2020 product design note. URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiCs-Hjx6_tAhUyy4UKHRLxCt8QFjAAegQIARAC&url=https%3A%2F%2Fwww.elia.be%2F-%2Fmedia%2Fproject%2Felia%2Felia-site%2Felectricity-market-and-system---document-library%2Fbalancing---balancing-services-and-bsp%2F2019%2F2019-design-note-mfrr-2020_version-01032019.pdf&usg=A0vVaw3xBQ_jIskuZDsY_GU6SA83.
- [ELIe] ELIA. Preliminary report on ELIA’s findings regarding the design of a scarcity pricing mechanism for implementation in Belgium. URL: https://www.elia.be/-/media/project/elia/elia-site/public-consultations/2020/20200930_elia_preliminary-report-scarcity-pricing_en.pdf.
- [ELIf] ELIA. Préparation de l’enchère crm y-4 pour la période de livraison 2025-26: Rapport du gestionnaire du réseau contenant des informations pour la détermination du volume à contracter et des propositions de paramètres

- spécifiques. URL: https://www.elia.be/-/media/project/elia/elia-site/users-group/crm-implementation/documents/20201204_dy2025---y-4-auction---calibration-report_fr.pdf.
- [ELIg] ELIA. The strategic reserve – a mechanism to cover structural shortages. URL: https://www.elia.be/-/media/project/elia/elia-site/strategic-reserve/722-adequacy_strategic-reserve/01_product-sheet/201810_sr_product-sheet_uk.pdf.
- [ELIh] ELIA. Terms and conditions for balancing service providers for manual frequency restoration reserve (mfrr). URL: https://www.elia.be/-/media/project/elia/elia-site/electricity-market-and-system---document-library/balancing---balancing-services-and-bsp/2020/20200203_bsp-contract-mfrr_en.pdf.
- [ELI18] ELIA. Study report on scarcity pricing in the context of the 2018 discretionary incentives, 2018.
- [ELI19] ELIA. Tariffs for maintaining and restoring the residual balance of individual access responsible parties 2020-2023. Technical report, Belgian Transmission System Operator, 2019. URL: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwj6-nhrcnuAhXl4IUKHfMJD8oQFjACegQIAxAC&url=https%3A%2F%2Fwww.elia.be%2F-%2Fmedia%2Fproject%2Felia%2Felia-site%2Fcustomers%2Ftariffs-and-invoicing%2Ftariffs-and-invoicing%2Fen%2Fgrille-tarifaire-2020-2023-onevenwicht-env1.pdf&usg=A0vVaw0LfoCj8YbDn7PPZqF-FdQy>.
- [EPEa] EPEXSPOT. EPEXSPOT and ECC successfully launch intraday auctions in Austria, Belgium, France and the Netherlands. URL: <https://www.epexspot.com/en/news/epex-spot-and-ecc-successfully-launch-intraday-auctions-austria-belgium-france-and-netherlands>.
- [EPEb] EPEXSPOT. New record volume traded on EPEXSPOT in 2020. URL: <https://www.epexspot.com/en/news/new-record-volume-traded-epex-spot-2020>.
- [EPEc] EPEXSPOT. Nominated electricity market operators (NEMOs) and transmission system operators (TSOs) are pleased to announce that XBID was successfully launched on tuesday 12th june. first deliveries were achieved on wednesday 13th June. URL: <https://www.epexspot.com/en/news/european-cross-border-intraday-xbid-solution-and-10-local-implementation-projects-successful>.
- [EPEd] EPEXSPOT. Press release: EPEXSPOT intraday markets reach all-time high in 2016. URL: https://www.epexspot.com/sites/default/files/download_center_files/2017-01-11_EPEX%20SPOT_2016_Annual%20Press%20Release.pdf.

- [EPEe] EPEXSPOT. The role and importance of power exchanges. URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjJ-eCW06rtAhXqvKQKHYZnBfUQFjAFegQICxAC&url=https%3A%2F%2Fwww.energy-community.org%2Fdam%2Fjcr%3A406a8587-566b-4126-b584-166ee5632050%2FECEBREG102018_EPEXSPOT.pdf&usg=A0vVaw20kKQvD2uUaBAZ9SM_4dId.
- [EPEf] EPEXSPOT. Trading brochure. URL: https://www.epexspot.com/sites/default/files/download_center_files/Trading%20Brochure.pdf.
- [EPEg] EPEXSPOT. Trading on EPEXSPOT 2020. URL: https://www.epexspot.com/sites/default/files/download_center_files/20-01-24_TradingBrochure.pdf.
- [ERC15] ERCOT. Ercot market training: Purpose of ORDC, methodology for implementing ORDC, settlement impacts for ORDC, 2015. URL: http://www.ercot.com/content/wcm/training_courses/107/ordc_workshop.pdf.
- [ES11] A. Ehrenmann and Y. Smeers. *Stochastic Equilibrium Models for Generation Capacity Expansion*, volume 163 of *Stochastic Optimization Methods in Finance and Energy, International Series in Operations Research and Management Science, Part 2*, pages 273–310. Springer, 2011.
- [ESFK18] E. Engmark, H. Sandven, S. Fleten, and G. Klaboe. Stochastic multistage bidding optimisation in an intraday market with limited liquidity. In *15th International Conference on the European Energy Market (EEM)*, 2018.
- [Eur] Eurostat. Share of renewable energy in gross final energy consumption. URL: https://ec.europa.eu/eurostat/databrowser/view/t2020_rd330/default/table?lang=en.
- [Eur17] European Commission. Commission regulation (EU) 2017/2195 of 23 november 2017 establishing a guideline on electricity balancing, 2017. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R2195&from=EN>.
- [Eur19] European Commission. Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 june 2019 on the internal market for electricity (recast). *Official Journal of the European Union*, 2019.
- [Gie19] P. Giesbertz. The power market design column - the scarcity of scarcity pricing, 2019. URL: <https://www.linkedin.com/pulse/power-market-design-column-scarcity-pricing-paul-giesbertz/>.
- [GM15] E. Garnier and R. Madlener. Balancing forecast errors in continuous-trade intraday markets. *Energy Systems*, 6(3):361–388, Sep. 2015.
- [Gol85] K. Golabi. Optimal inventory policies when ordering prices are random. *Operations Research*, 33(3), 1985.

- [HDV16] H. Hoschle and K. De-Vos. Implementation of a strategic reserve in Belgium: Product design and market results. In *Proceedings of the International Council on Large Electric Systems (CIGRE)*, 2016.
- [Hog05] W. Hogan. On an ‘energy only’ electricity market design for resource adequacy. Technical report, Center for Business and Government, JFK School of Government, Harvard University, September 2005.
- [Hog13] W. Hogan. Electricity scarcity pricing through operating reserves. *Economics of Energy and Environmental Policy*, 2(2):65–86, 2013.
- [HP19] W. Hogan and S. Pope. PJM reserve markets: Operating reserve demand curve enhancements. Technical report, Harvard University, 2019. URL: https://sites.hks.harvard.edu/fs/whogan/Hogan_Pope_PJM_Report_032119.pdf.
- [IRE21] IRENA. Utility-scale batteries innovation landscape brief, 2021. URL: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2019/Sep/IRENA_Utility-scale-batteries_2019.pdf.
- [KEF15] G. Klaboe, A. Eriksrud, and S. Fleten. Benchmarking time series based forecasting models for electricity balancing market prices. *Energy Systems*, 6(1):43–61, Mar. 2015.
- [Kie17] R. Kiesel. Modeling market order arrivals on the intraday power market for deliveries in Germany with Hawkes processes with parametric kernels. In *Energy Finance Christmas Workshop*, 2017.
- [Kin69] B. Kingsman. Commodity purchasing. *Journal of the Operational Research Society*, 20(1):59–79, Mar. 1969.
- [KKP20] M. Kremer, R. Kiesel, and F. Paraschiv. Intraday electricity pricing of night contracts. *Energies*, 13(17):4501, Sep. 2020.
- [KP17] R. Kiesel and F. Paraschiv. Econometric analysis of 15-minute intraday electricity prices. *Energy Economics*, 64:77–90, May 2017.
- [KSW13] M. Khodayar, M. Shahidehpour, and L. Wu. Enhancing the dispatchability of variable wind generation by coordination with pumped-storage hydro units in stochastic power systems. *IEEE Transactions on Power Systems*, 28(3):2808–2818, Aug. 2013.
- [Lin61] D. Lindley. Dynamic programming and decision theory. *Applied Statistics*, 10(1):39–52, 1961.
- [LSB18] D. Lee, H. Shin, and R. Baldick. Bivariate probabilistic wind power and real-time price forecasting and their applications to wind power bidding strategy development. *IEEE Transactions on Power Systems*, 33(6):6087–6097, Nov. 2018.
- [LW21] N. Lohndorf and D. Wozabal. Gas storage valuation in incomplete markets. *European Journal of Operational Research*, 288(1):318–330, Jan. 2021.

- [MO18] H. Martin and S. Otterson. German intraday electricity market analysis and modeling based on the limit order book. In *15th International Conference on the European Energy Market (EEM)*, 2018.
- [Mor59] W. Morris. Some analysis of purchasing policy. *Management Science*, 5(4):443–452, Jul. 1959.
- [MRRFJC16] C. Monteiro, I. Ramirez-Rosado, L. Fernandez-Jimenez, and P. Conde. Short-term price forecasting models based on artificial neural networks for intraday sessions in the Iberian electricity market. *Energies*, 9(9), 2016.
- [MS06] J. Matevosyan and L. Soder. Minimization of imbalance costs trading wind power on the short-term power market. *IEEE Transaction on Power Systems*, 21(3):1396–1404, Aug. 2006.
- [ND07] V. Nanduri and T. Das. A reinforcement learning model to assess market power under auction-based energy pricing. *IEEE Transaction on Power Systems*, 22(1):85–95, Feb. 2007.
- [NRE03] NREL. Hydropower setting a course for our energy future, 2003. URL: <https://www.nrel.gov/docs/fy04osti/34916.pdf>.
- [OF15] C. O’Dwyer and D. Flynn. Using energy storage to manage high net load variability at sub-hourly time-scales. *IEEE Transactions on Power Systems*, 30(4):2139–2148, Jul. 2015.
- [Pap] A. Papavasiliou. An overview of probabilistic dimensioning and scarcity pricing with a focus on the greek electricity market. URL: <https://www.rae.gr/wp-content/uploads/2021/05/Report-II-ProbabilisticDimensioning-final.pdf>.
- [Pap20] A. Papavasiliou. Scarcity pricing and the missing European market for real-time reserve capacity. *The Electricity Journal*, 2020.
- [PB21] A. Papavasiliou and G. Bertrand. Market design options for scarcity pricing in european balancing markets. *IEEE Transactions on Power Systems*, 2021.
- [PM16a] W. Powell and S. Meisel. Tutorial on stochastic optimization in energy-part ii: An energy storage illustration. *IEEE Transactions on Power Systems*, 31(2):1468–1475, Mar. 2016.
- [PM16b] W. Powell and S. Meisel. Tutorial on stochastic optimization in energy-part i: modeling and policies. *IEEE Transactions on Power Systems*, 31(2):1459–1467, Mar. 2016.
- [PP91] M. Pereira and L. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.
- [PS] A. Papavasiliou and Y. Smeers. Remuneration of power generation capacity in conditions of scarcity in Belgium. URL: <https://ap-rg.eu/wp-content/uploads/2020/07/CEER2016.pdf>.

- [PS17] A. Papavasiliou and Y. Smeers. Remuneration of flexibility using operating reserve demand curves: A case study of Belgium. *The Energy Journal*, 38(6):105–135, 2017.
- [PSB18] A. Papavasiliou, Y. Smeers, and G. Bertrand. An extended analysis on the remuneration of capacity under scarcity conditions. *Economics of Energy and Environmental Policy*, 7(2), 2018.
- [PSd19] A. Papavasiliou, Y. Smeers, and G. de Maere d’Aertrycke. Study on the general design of a mechanism for the remuneration of reserves in scarcity situations. Technical report, UCLouvain, 2019. URL: <https://www.creg.be/sites/default/files/assets/Publications/Notes/Z1986Annex.pdf>.
- [PSdMd20] A. Papavasiliou, Y. Smeers, and Gauthier de Maere-d’Aertrycke. Market design considerations for scarcity pricing: A stochastic equilibrium framework. *The Energy Journal*, 2020.
- [RAE21] RAE. Public consultation – cost of new entry, 2021. URL: <https://www.rae.gr/2021/05/28/public-consultation-cost-of-new-entry/>.
- [RAP16] P. Gruet R. Aid and H. Pham. An optimal trading problem in intraday electricity markets. *Mathematics and Financial Economics*, 10(1):49–85, Jan. 2016.
- [RS15] D. Ralph and Y. Smeers. Risk trading and endogenous probabilities in investment equilibria. *SIAM Journal on Optimization*, 25(4):2589–2611, 2015.
- [SB18] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. MIT press, 2018.
- [Sch17] D. Schiro. Flexibility procurement and reimbursement: A multiperiod pricing approach. In *FERC Technical Conference*, 2017.
- [SEM15] A. Skajaa, K. Edlund, and J. Morales. Intraday trading of wind energy. *IEEE Transactions on Power Systems*, 30(6):3181–3189, Nov. 2015.
- [Sto02] S. Stoft. *Power System Economics: Designing Markets for Electricity*. IEEE Press and Wiley Interscience, 2002.
- [SZ19] R. Steinhert and R. Ziel. Short-to mid-term day-ahead electricity price forecasting using futures. *The Energy Journal*, 40(1):105–128, 2019.
- [TEN] TENNET. Balancing group contract. URL: https://www.tennet.eu/fileadmin/user_upload/The_Electricity_Market/German_Market/Grid_customers/contracts/BNetzA-BKC_englisch.pdf.
- [Tes21] Tesla. Powerwall, 2021. URL: https://www.tesla.com/sites/default/files/pdfs/powerwall/Powerwall%20AC_Datasheet_en_northamerica.pdf.

- [TR99] J. Tsitsiklis and B. Van Roy. Optimal stopping of markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Transactions On Automatic Control*, 44(10), Oct. 1999.
- [US 21a] US Department Of Energy. U.S. hydropower market report, 2021. URL: <https://www.energy.gov/sites/prod/files/2021/01/f82/us-hydropower-market-report-full-2021.pdf>.
- [US 21b] US Energy Information administration. Utility-scale batteries and pumped storage return about 80% of the electricity they store., 2021. URL: <https://www.eia.gov/todayinenergy/detail.php?id=46756>.
- [WD92] C. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, May 1992.
- [Wil92] R. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4):229–256, May 1992.
- [YLP10] N. Yu, C. Liu, and J. Price. Evaluation of market rules using a multi-agent system method. *IEEE Transaction on Power Systems*, 25(1):470–479, Feb. 2010.
- [YQLS19] Y. Ye, D. Qiu, J. Li, and G. Strbac. Multi-period and multi-spatial equilibrium analysis in imperfect electricity markets: A novel multi-agent deep reinforcement learning approach. *IEEE Access*, 7:130515 – 130529, Sep. 2019.
- [YQS⁺20] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos, and G. Strbac. Deep reinforcement learning for strategic bidding in electricity markets. *IEEE Transactions on Smart Grid*, 11(2):1343–1355, Mar. 2020.
- [Zie17] F. Ziel. Modeling the impact of wind and solar power forecasting errors on intraday electricity prices. In *14th International Conference on the European Energy Market (EEM)*, 2017.
- [ZT20] M. Ziegler and J. Trancik. Re-examining rates of lithium-ion battery technology improvement and cost decline. *Energy & Environmental Science*, 2020. URL: <https://www.nrel.gov/docs/fy04osti/34916.pdf>.