Université catholique de Louvain
École Polytechnique de Louvain

Center for Operations Research and Econometrics

Real-Time Pricing in Integrated European Electricity Markets

Jacques Cartuyvels

Thesis submitted in partial fulfillment of the requirements for the degree of
Docteur en sciences de l'ingénieur et technologie

Supervisor:
Anthony Papavasiliou
(NTUA, Greece)
Per Agrell
(UCLouvain, Belgium)

Jury:
Philippe Chevalier (UCLouvain, Belgium)
Bert Willems (UCLouvain, Belgium)
Derek Bunn (London Business School, United-Kingdom)

Chair:
Philippe Chevalier (UCLouvain, Belgium)

PhD Organization

Jacques Cartuyvels
UCLouvain
École Polytechnique de Louvain
Center for Operations Research and Econometrics

Thesis Supervisor

Anthony Papavasiliou
Assistant Professor, National Technical University of Athens
Department of Electrical and Computer Engineering

Per Agrell
Professor, UCLouvain
Louvain School of Management
Center for Operations Research and Econometrics

Supervisory Committee

Yves Smeers
Professor, UCLouvain

Gauthier de Maere d'Aertrycke
Director, Engie

Abstract

This thesis provides practical answers to specific market design issues related to the short-run pricing of electricity in the context of the integration of European balancing electricity markets and the introduction of scarcity pricing through an operating reserve demand curve (ORDC). Scarcity pricing refers to the practice of pricing electricity above the marginal cost of the marginal generator. It is an essential component of electricity market design to reach the efficient long-term equilibrium as it allows plant owners to recuperate their investment cost.

Chapter 2 discusses the parametrization of the ORDC. The introduction of an ORDC is an administrative measure warranted by inadequate scarcity pricing caused by a lack of a strong responsive demand side. The administrative nature of the mechanism necessitates a tool to assess its parametrization. This chapter presents a short-term system operation simulator of Belgium to validate the calibration of ORDCs. The simulator measures the trade-off between the cost of operation and system reliability resulting from the ORDC calibration. Eight variants of ORDCs are considered, and this analysis serves as the basis for supporting a recommendation to the Belgian regulatory authority for the roll-out of scarcity pricing in Belgium [CREG, 2021].

Chapter 3 analyses how scarcity pricing can be implemented in the European setting. It is a bridge between the scarcity pricing discussion and the European balancing market discussion. It investigates the ambiguity stemming from the introduction of scarcity adders and the potential discrepancy between the balancing price used to remunerate flexibility provider supplying balancing energy and flexibility consumers generating imbalances. Three options for introducing scarcity adders are discussed: (i) an adder on the imbalance price, (ii) an adder on the imbalance and balancing price, and (iii) an adder on the imbalance and balancing price and the introduction of a real-time market for reserve. This analysis highlights the detrimental effect of unilaterally introducing adders with a real-time market for reserve in an integrated European setting. These adders induce inefficient out-of-merit dispatch.

Chapter 4 presents another source of discrepancy between the balancing and imbalance prices: the multiplicity of balancing energy product. Balancing the market is a continuous process that involves the activation of automatic and manual frequency regulation reserve (aFRR and mFRR). These products are

traded on different time-scales and their price are used to form the imbalance price. This chapter examines the impact of (i) the imbalance pricing scheme, (ii) the system operator's mFRR activation strategy, and (iii) the balancing capacity demand curve on the balancing market equilibria.

Acknowledgements

This thesis marks the end of my PhD journey. I am incredibly grateful to the people who accompanied me during those five years.

First and foremost, I wish to thank Anthony, my supervisor. We have been working together for six years, and I could not have hoped for a better mentor and friend. I am confident that I would not have been able to complete my PhD if it wasn't for your support and kindness.

My gratitude also goes to Gilles. I really enjoyed our collaboration over the last four years and our weekly meetings. The dynamic of our research trio was one of the best part of my PhD.

I would also like to thank my supervisory committee, Yves Smeers and Gauthier de Maere d'Aertrycke, and my defense committee, Prof. Per Agrell, Prof. Philippe Chevalier, Prof. Bert Willems, and Prof. Derek Bunn, for their support and insights. I am especially grateful to my badminton partner, Bert, for his challenging discussions in the CORE lounge.

I also wish to thank the CORE administrative staff, Catherine, Marie and Margaux. Your efforts to provide me with chocolate will not be forgotten.

I will not forget the friends I made at CORE over the years. I have fond memories with the energy group, Ilyes, Céline, Daniel, Quentin, Jehum, and Nicolas, and the rest of the CORE buddies, Antoine, Martial, Thomas, Robin, Hakime, and all the others.

Lastly, I am grateful for the support of my family and for the love of my wife.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Climate Change and the Need for Decarbonization

Climate change is the greatest challenge of our generation, and its effects are already visible. Global temperature has increased by 1.1 degrees Celsius during the period 2011-2020 compared to the period 1850-1900 [IPCC, 2023]. This has resulted in heatwaves, human fatalities, and an increase in wildfire frequency and intensity, but this is only the tip of the iceberg. Other extreme climatic events such as droughts and floods are endangering the food security of millions of people and generating political instability. Additionally, global warming increases the risk of reaching a climate tipping point with escalating effects that are difficult to predict.

There is a scientific consensus that this crisis has been caused by human activity and the emission of *greenhouse gases* (GHGs) and that only a drastic reduction in GHG emissions would limit the temperature increase to below 1.5°C or 2°C, as shown in figure 1.1 from the *Intergovernmental Panel on Climate Change* (IPCC) report [IPCC, 2023]. These targets have been set at the COP21 in Paris in 2015. On a European level, the *fit for 55* package has passed laws to reach climate neutrality by 2050 and a 55% decrease in greenhouse gas emissions in 2030 compared to 1990. The objective in terms of renewable is to reach 42.5-45% of overall energy sources at EU level by 2030.

The main greenhouse gases are *carbon dioxide* (CO_2), which is emitted when burning fossil fuel, and *methane*, which is emitted by livestock. Both emission sources must be reduced to reach the climate target. In particular, net global CO_2 emissions need to be reduced to zero by 2050. There exist different pathways to decarbonize our industry, transport, and energy generation to achieve the *energy transition* but they all share a common feature: the *electrification* of carbon-intensive processes. The advantage of electricity as an energy vector is that it can be produced carbon-free or with limited carbon dioxide emission. Transport can be decarbonized by electrifying the car and bus fleets. Heating

can be electrified with heat pumps. Some industry processes can also be electrified. Processes that cannot be electrified due to technical constraints can often be modified to use *green hydrogen* that is produced with renewable electricity.

The energy transition will lead to an increased reliance on electricity. Electricity demand will rise and, at the same time, classic fossil-fuel-based electricity generators will need to be decommissioned. A key component for the success of the energy transition is the design of the markets that are used for trading electricity. Their designs should ensure the most efficient allocation of resources to support the large-scale integration of renewable electricity that is envisioned by the energy transition.

1.2 Electricity Market Design under Deep Renewable Integration

Electricity systems are designed to transmit electric power from generation assets to loads through the electricity network. These systems have evolved from regulated vertically integrated monopolies to competitive *electricity markets* and are now on the verge of their second revolution due to the energy transition. They have always aimed at providing *reliable* and *affordable* electricity, but the need for decarbonization has introduced a new objective: *sustainability*. This objective is fulfilled by integrating renewable energy sources (RES), that are characterized by their intermittency and near-zero marginal cost. The integration of RES and the departure from dispatchable thermal power plants has triggered a debate on the ability of modern electricity markets to support the energy transition and to maintain the efficiency and reliability objectives.

Before the 1990s and the electricity liberalization, electricity systems were operated by regulated vertically integrated monopolies. State-backed utilities operated the entirety of electricity generation, high-voltage transmission, low-voltage distribution, and retail. Responsibilities could be split between local distribution networks and retail, and regional transmission networks and generation assets, but conceptually these systems were run by a single central entity [Cramton, 2017]. The main objective of the central operator was to provide reliable electricity by ensuring the feasibility of the physical constraints governing the exchange of electricity. The first such constraint is that electricity supply and consumption should be balanced at all times under the transmission and distribution grid constraints. The second constraint is that the capacity installed on the network should be sufficient to prevent load curtailment. These two components of "reliability" are referred to as *security* and *adequacy* respectively. The usual efficiency drawbacks afflicting regulated monopolies [Joskow, 2019] eventually led to the creation of electricity markets and the *unbundling* of electricity systems.

The unbundling of the electricity sector refers to the segmentation of the vertically integrated monopolies into generation companies (GENCOs) owning and operating power plants, retailers supplying energy to small local con-

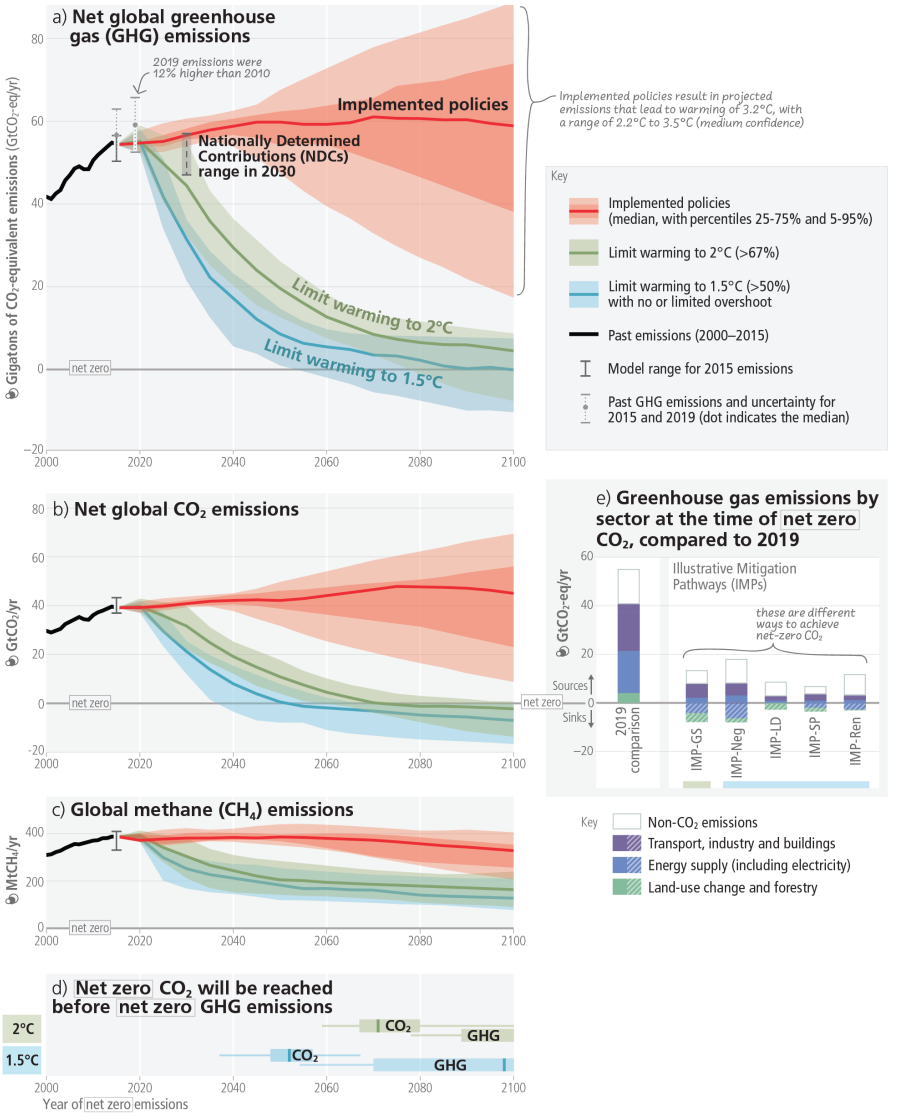


Figure 1.1: Limiting warming to 1.5°C and 2°C involves rapid, deep and in most cases immediate greenhouse gas emission reductions (from [IPCC, 2023]).

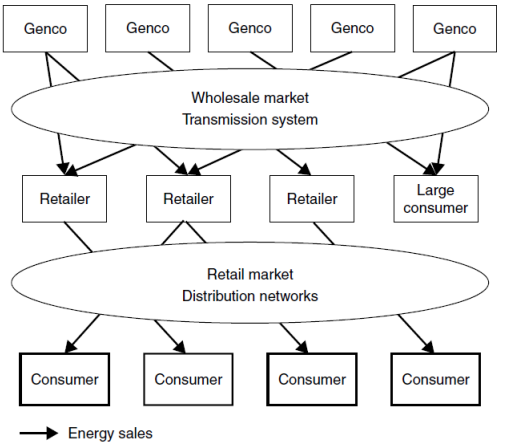


Figure 1.2: Structure of liberalized electricity markets (from [Kirschen and Strbac, 2018]).

sumers, a transmission system operated by a *system operator*, and local distribution networks. The natural division is now anchored around the split between *wholesale* and *retail market*. Gencos and retailers compete with *large consumers* connected to the transmission grid in the wholesale market whereas *small consumers* connected to distribution networks purchase electricity from retailers on the retail market, as illustrated in figure 1.2 [Kirschen and Strbac, 2018]. Electricity markets in this manuscript refer to wholesale markets and the intricacies of the distribution networks and retail markets are ignored.

Electricity markets are institutions that aim at ensuring an efficient allocation of resources in the *short-* and *long-run*. In the short-run, they should allow for the scheduling and dispatch of power plants that maximize social welfare, and, in the long-run, they should signal the economical entry and exit of capacity [Joskow, 2019]. More specifically, short-term market signals, the electricity spot prices, should align the short-term scheduling and dispatch incentives of private agents with the optimal short-term allocation of a benevolent system operator. Similarly, long-term market signals, the expected revenues from the electricity spot prices, should support the optimal capacity mix generated by a central planner not subject to the inefficiencies stemming from being a regulated monopoly. In theory, ideal competitive electricity markets can achieve this equilibrium. In practice, market failures, and the current implementation of electricity markets, prevent this idealized outcome.

One of the main objectives of liberalization was to produce incentives for efficient capacity investment. The central planner approach had led to over-investment as there were no incentives to control cost, and the alternative, curtailment of electricity due to a lack of capacity, was deemed serious enough, both economically and politically, to justify expensive investment. Unfortu-

nately, modern competitive markets have failed to provide sufficient long-term signals for investment and the opposite effect has been observed. In various regions of the world, spot prices alone have been deemed insufficient to induce investment in additional capacity. Reaching the efficient long-term equilibrium has been particularly challenging due to the so-called *missing money* problem that stems from inadequate pricing during periods of scarcity. This market failure has been exacerbated by the energy transition and the large-scale integration of renewable resources that are subsidized outside of the market. The introduction of a large fleet of assets with low marginal cost has generated a *merit order effect* in which more expensive dispatchable power plants are pushed further away in the activation order. This reduces their expected revenues from the spot market and hinders investment in new capacity.

The integration of RES and the induced merit order effect does not prevent sustained high prices event such as the one caused by the 2022 European gas crisis caused by the disruption of the European gas supply chain. This has caused an increase in natural gas and oil prices that affected the immediate evolution of the European energy market. The financial repercussion of this crisis has been massive for electricity retailers, while certain technologies on the supply side of the market are enjoying comfortable profit margins. However, the electricity price increase does not necessarily translate to a reduction of the missing money. Peaking unit that were suffering from insufficient profit margin are often gas power-plant. In the absence of market manipulation, their increased revenues should be countered by an increased production cost limiting their profit margin.

The discussion on how to reach an efficient long-term equilibrium is central to the success of the energy transition. Market design for decarbonized electricity systems needs to support a long-term welfare-maximizing equilibrium that accounts for reliability and sustainability constraints. There are divergent perspectives on how to best reach this objective. One school of thought advocates for stronger and improved energy spot prices [Hogan, 2019, Cramton, 2017, Mays and Jenkins, 2023]. The arguement here is that improved energy pricing and adequate financial contracts can be sufficient for addressing the missing money problem and restore investment incentives. Another school of thought contends that short-term operation and long-term investment are now largely disconnected [Newbery et al., 2018, Joskow, 2019]. Short-term markets are still necessary for the optimal dispatch of assets and balancing the system, however investments in new capacity should follow a central planner mandates through capacity remuneration mechanisms in which private investors bid their missing money.

Regardless of the strategy for attaining a long-term equilibrium, there is a consensus that improving the pricing of short-run electricity will be necessary to cope with the intermittency brought by the integration of RES. This intermittency poses a challenge to the balance of electricity supply and demand. It has the potential to compromise the reliability of electricity systems. It can be addressed by either shifting the demand and supply over space via interconnec-

tors, by shifting the demand and supply over time via storage, or by resorting to flexible demand or supply [Newbery et al., 2018]. This last option can be accessed by improving the pricing schemes that remunerate flexibility. Flexibility for the supply side is composed of responsive power plants that can modify their dispatch on demand. Flexibility on the demand side is called *demand response* and relies on residential, industrial, and commercial consumers who are willing to modify their consumption schedule as a function of the needs of the system. Flexible capacity can be categorized based on the activation time, ranging from a few minutes to a few hours, and the direction, downward activation for compensating positive imbalance when the system is long and upward activation for compensating negative imbalance when the system is short. The total need for upward and downward flexibility in Belgium has been estimated to increase from 4320 MW to 7380 MW and from 3500 MW to 5960 MW by 2034 [ELIA, 2024]. These figures represent a 70% increase in flexible capacity and show the need for efficient short-run incentives.

This thesis is thus part of the literature in electricity market design on pricing in short-term electricity markets to support the energy transition and the large-scale integration of RES. It provides practical answers to specific policy issues related to market design features that may hinder short-run price formation. This thesis focuses on the short-term equilibrium and ignores the long-run issues with the assumption that improving the short-run can only be beneficial to the long-run equilibrium.

These market features are studied in the context of the integration of European electricity markets. More specifically, this thesis contributes to the literature on the integration of *balancing markets* in the EU. European balancing markets are held in real time in order to ensure the continuous balance of electricity consumption and generation. They are the successor of power system mechanisms conducted by the system operators, and consequently, have been designed with a focus on operational priorities rather than economics principles. This thesis aims at rationalizing their designs based on economic fundamentals.

The second focal point of this analysis is the use of *operating reserve demand curves* (ORDCs) for implementing *scarcity pricing*. Scarcity pricing refers to the practive of allowing prices to rise above the marginal cost of production of the marginal generator in times of scarcity. Scarcity pricing is a necessary component for reaching the efficient long-term equilibrium as it allows plant owners to recuperate their investment costs. The use of ORDCs for scarcity pricing was proposed by Stoft [Stoft, 2002] and formalized by Hogan [Hogan, 2005]. Scarcity pricing has commonly been associated with US-style electricity markets and this thesis investigates how this mechanism can be adapted to European-style markets.

1.3 European Balancing Markets

Balancing the market is the last step in the sequence of electricity markets. It refers to the real-time dispatch of electricity assets to balance electricity generation and consumption. The efficiency of this process is a critical element for the success of the energy transition, as it ensures the reliability of the electricity system. Equally critical, this process sets the real-time price that drives the wholesale electricity price in the day-ahead and other forward markets [Cervigni and Perekhodtsev, 2013].

The balancing process in Europe is organized by *transmission system operators* (TSOs). They procure *balancing energy* to compensate for the real-time *imbalance* that they measure on their transmission grid. They act as intermediaries between flexibility suppliers providing balancing energy and flexibility consumers generating imbalances. The European terminology for flexibility suppliers and flexibility consumers are *balancing service providers* (BSPs) and *balancing responsible parties* (BRPs) respectively. Flexibility suppliers and consumers can be on both sides of the traditional dichotomy between energy demand and supply. Energy suppliers, such as gas power plants, can adjust their scheduling dispatch to provide balancing energy but wind turbines can generate imbalances due to inaccurate wind forecasts. Similarly, loads can generate imbalances when they consume more or less than their agreed consumption schedule but demand response agents can provide balancing energy by modulating their real-time consumption. This shows that differences between balancing energy and imbalances are more a matter of terminology rather than a physical distinction. The need for an entity to coordinate the balancing action is warranted by the speed of adjustment required to maintain the stability of the system. Prices alone are not sufficient [Wilson, 2002]. Flexibility in this context refers to short-term capacity that can adjust its production or consumption schedule. This does not include the long-term flexibility that can be used for compensating for seasonal discrepancies between production and generation.

Balancing markets refer to the real-time *balancing energy* auctions for the activation of flexible assets but they are not the only components of the balancing process. The balancing energy auctions are preceded by *balancing capacity* auctions and followed by the *imbalance settlement*. Balancing capacity is the European terminology for *reserve*. It refers to the generating capacity available to the system operator to react to the short-term uncertainty of electricity production and consumption and to contingency event such as the failure of transmission lines or power plants. These reserves are procured to ensure an adequate supply of balancing energy in real time. The tendered quantity that is required to ensure the reliability of the system is determined through a *reserve dimensioning* process. Balancing capacity is a *public good* financed by all the grid users through the grid tariffs. Imbalance settlement refers to the pricing of imbalances. Flexibility consumers are charged ex-post for the imbalance they generate. Imbalances are settled at the *imbalance price* which is determined

by the activation cost of balancing energy [Brijs et al., 2017].

1.3.1 US-Style Real-Time Markets

European and US policy makers have designed their electricity markets differently. US-style markets, such as PJM, MISO and ERCOT, typically rely on *integrated markets* in which the scheduling is determined by *unit commitment* problems and the dispatch by *economic dispatch* problems run by *independent system operators* (ISOs). Power plants participate individually in these optimization problems with explicit technical constraints. In contrast, European-style markets typically rely on *exchange-based markets* in which power-producing companies submit price-quantity bids representing their willingness to produce to a *day-ahead* auction [Cramton, 2017]. There are recourse auctions called *intraday markets* but the production schedules are finalized between one hour and fifteen minutes before electricity delivery. This leaves some time for European system operators to operate the balancing markets. The greater coordination brought by the integrated approach allows US system operators to rely on the *real-time* markets to set the final production schedule. This schedule is updated every five minutes and the uncertainty adjustment within five minutes interval are handled by *automatic generation control* (AGC) priced outside of the wholesale market. Arbitrary deviations from the production schedules are often settled at a premium called *uninstructed imbalance penalty* or *uninstructed deviation penalty*.

US-style real-time markets can be represented as a welfare maximization problem where a system operator aims at maximizing the benefit of holding reserve minus the cost of covering the inelastic imbalance given power plants (PP) with a given marginal cost (C) and maximum technical production (Q), given real-time demand (D) and given the marginal valuation for reserve by the TSO ($MBR(\cdot)$). This exposition ignores the indivisibilities and ramp constraints due to the operational characteristics of the plants and assumes linear production costs to focus on the market structure. The power plant capacity is split between energy (p) and reserve (r) (or balancing energy and balancing capacity in EU terminology), and the duality conditions of the optimization problem unambiguously characterize the remuneration of both products. The model, in stylized form, can be expressed as follows:

$$\max_{dr,r,p} \quad \int_0^{dr} MBR(x)dx - \sum_{g \in PP} C_g \cdot p_g \tag{1.1a}$$

$$s.t. \quad (\lambda): \quad \sum_{g \in PP} p_g = D \tag{1.1b}$$

$$(\lambda^R): \quad \sum_{g \in PP} r_g = dr \tag{1.1c}$$

$$(\mu_g): \quad p_g + r_g \leq Q_g \qquad \forall g \in PP \tag{1.1d}$$

$$p, r, dr \geq 0 \tag{1.1e}$$

Energy is remunerated at the energy price, which is the dual variable related to the market clearing constraint for energy (λ). Reserve capacity is remunerated at the reserve price, which is the dual variable related to the market clearing constraint for reserve (λ^R). Note that constraint (1.1d) results in a coupling of the energy and reserve prices and that dr represents the remaining balancing capacity in the system, which is also the satisfied demand for reserve.

An inspection of the KKT conditions related to the energy production and reserve variables and the remaining balancing capacity variable connects the price for energy and reserve to the remaining capacity in the system:

$$0 \leq r_g \perp \mu_g - \lambda^R \geq 0 \tag{1.2a}$$
$$0 \leq p_g \perp C_g + \mu_g - \lambda \geq 0 \tag{1.2b}$$
$$0 \leq dr \perp \lambda^R - MBR(dr) \geq 0 \tag{1.2c}$$

Constraint (1.2c) links the reserve price to the value of the marginal benefit function at the level of remaining balancing capacity in the system. Let us refer to a marginal generator as a generator that supplies both energy and reserve at the optimal solution, and let us index this generator by g'. Constraints (1.2b) and (1.2a) show that the scarcity rent for a marginal generator g' is equal to $\mu_{g'}$ and is further equal to the reserve price. In this stylized model without ramp constraints, $\mu_{g'}$ is equal to λ^R, and the energy price is then equal to the marginal cost of the marginal generator $C_{g'}$ augmented by the reserve price λ^R. These conditions ensure that a marginal generator is indifferent between supplying reserve or energy.

This welfare maximization problem is an energy and reserve economic dispatch. The system operator co-optimizes the procurement of both products. Co-optimization allows *scarcity pricing* to emerge naturally, with the price of energy rising above the marginal cost of the marginal generator, when the marginal value of an additional MWh of reserve is greater than zero. This feature is discussed more at length in the next section of the chapter on scarcity pricing.

In practice, balancing markets in the EU differ from the ideal co-optimization model presented in (1.1). This is the case even if we ignore the distinction between portfolio-based versus unit-based designs. A quantitative model of the European balancing market can be described as follows for flexibility suppliers (FS) submitting at a given marginal cost (C) and volume (Q) and flexibility consumers (FC) generating imbalances (IMB):

$$\max_{p} \quad -\sum_{g \in FS} C_g \cdot p_g \tag{1.3}$$
$$s.t. \quad (\lambda): \quad \sum_{g \in FS} p_g = \sum_{l \in FC} IMB_l \tag{1.4}$$
$$(\mu_i): \quad p_g \leq Q_g \qquad \forall g \in FS \tag{1.5}$$

$$p \geq 0 \tag{1.6}$$

The first difference with the US-style model is that TSOs do not value balancing capacity in real time and run an "energy-only"[1] optimization problem. This absence of a market for real-time balancing capacity raises issues for the forward reservation of balancing capacity, as already pointed out in [Papavasiliou et al., 2021]. Without a real-time valuation for reserve, the only opportunity cost that can be back-propagated to the day-ahead balancing capacity auction would arise from the payoff difference between the balancing market and the day-ahead spot market or from fixed cost effects.

The second point of differentation with US markets is that the law of one price [Jevons, 1871] is not respected for the trading of balancing energy. The balancing energy of flexibility suppliers (on the left-hand side of constraint (1.4)) is remunerated at the balancing price, whereas flexibility consumers (on the right-hand side of constraint (1.4)) face an imbalance price that is based on the balancing price but may differ from it. In Belgium, for example, an "alpha" component was introduced on top of the balancing price. The stated objective of such an adder is to incentivize flexibility suppliers to keep their imbalance low. The point of view of ELIA, the national transmission system operator of Belgium, is that the difference in pricing between imbalances caused by flexibility consumers and balancing energy provided by flexibility suppliers can be justified by the different goals of the price signals:

> *"The imbalance price incentivizes BRPs (flexibility consumers) to keep and/or restore system balance of their imbalance price area in accordance with the Electricity Balancing Regulation, while the balancing energy price reflects the price of the marginal bid selected in the uncongested area by the activation optimization function of the EU balancing platform"* [ELIA, 2021b].

In practice, flexible assets can move freely between active imbalance and participation in the balancing market, giving rise to arbitrage opportunities.

1.3.2 Multiple Balancing Products and Imbalance Pricing

The previous exposition assumes a single balancing product for covering imbalance. In practice, balancing the market is a continuous process that involves the activation of a variety of balancing products with different technical characteristics and activation times. These products are tailored to specific balancing issues. In European balancing markets, there are three main products: *frequency containment reserve* (FCR), *automatic frequency restoration reserve*

[1] Beware that the term "energy-only" can refer to markets without operating reserve in the dispatch or markets not relying on capacity markets for reaching the long-term equilibrium depending on the context.

(aFRR) and *manual frequency restoration reserve* (mFRR). These products were also referred to in the past as primary, secondary, and tertiary reserves.

FCR reacts proportionally and automatically to the frequency of electricity. This product handles transient frequency deviations. FCR is often traded as a purely capacity product. There is no energy component to remunerate the activation of FCR as the symmetry of upward and downward activation is assumed to even out. aFRR is used to relieve FCR capacity. It reacts almost instantaneously to larger imbalances and is traded on a four-second basis with a full activation time of five to seven and a half minutes. The activation of aFRR is based on an automatic controller. Alternatively, mFRR can be used to compensate for larger imbalances. mFRR activation is performed manually on a fifteen-minute basis.

The US reserve products vary from one ISO to another and the product definitions differ from the standardized European ones. The main difference in products concerns secondary reserve (aFRR). It is controlled and remunerated through an auxiliary market that is distinct from the real-time market. There also exist longer-lasting reserves that can be activated during the integrated real-time dispatch. They are closer to mFRR in terms of characteristics. They are often categorized as *contingency reserve* and can be further divided into *spinning* and *non-spinning* reserve, depending on whether the assets providing the reserve are connected to the grid.

The multiplicity of real-time energy products in European balancing energy markets, aFRR and mFRR, raises an issue concerning the pricing of imbalances that is absent in US-style markets. There is no ambiguity in the real-time energy prices generated by an integrated economic dispatch model and uninstructed deviation are settled outside of the real-time market. On the other hand, it is unclear how imbalance prices set on a fifteen-minute basis can incorporate both an aFRR and an mFRR price component, given the difference in timing granularity.

1.3.3 European Balancing Market Integration and Cross-Border Balancing Platforms

The next phase of the European electricity market integration relates to balancing markets. It follows the integration of day-ahead markets through the market coupling algorithm *Euphemia*. This algorithm connects the electricity price in 26 European countries and allocates transmission capacity to maximize welfare. It is considered the primary market for electricity in Europe. *Intraday* markets used for revising production and consumption schedules closer to physical delivery have also been integrated through the *single intraday coupling* platform.

One of the first steps in the integration of the European balancing market was to standardize the balancing products [Meeus, 2020]. National balancing markets in Europe have been characterised by specific features in terms of (i) traded balancing products, (ii) balancing capacity procurement (market-

based or mandatory contribution), settlement methods (regulated price, pay-as-bid, or pay-as-cleared) and cost recovery schemes (passed on to the customer through the grid tariff or covered by flexibility consumers causing imbalances), (iii) balancing energy activation (pro-rata or based on a merit-order), settlement methods and cost recovery schemes, (iv) the *imbalance settlement period* duration (from 15 minutes to 1 hour) and (v) imbalance price formation. These rules reflect the approach of diverse TSOs for covering imbalances and are a legacy of a time with less coordination between European member states in real time. The standardization process has been followed by several initiatives to allow for the cross-border trading of balancing products. Currently, (i) FCR capacity is jointly procured by the Austrian, Belgian, Czech, Danish, Dutch, French, German, Slovenian and Swiss TSOs, (ii) mFRR balancing energy activation is coordinated by the balancing platform MARI in Germany, Austria, and the Czech Republic, and (iii) the activation of aFRR balancing energy is optimized by the balancing platform PICASSO in Germany, Austria, the Czech Republic, Slovenia and Italy[2]. To be more specific on aFRR and mFRR, TSOs connected to the MARI platform submit a demand for mFRR balancing energy over the next fifteen minutes, potentially in the form of a demand curve, and the platform selects *balancing energy bids* at least cost in order to satisfy the demand given the network topology. Then, independent PICASSO runs are solved every four seconds based on the zonal imbalances that are yet to be covered. mFRR and aFRR balancing energy activated by the platforms are remunerated at the aFRR and mFRR platform prices respectively. The rules for setting the platform prices are framed by decision 2020/01 of ACER and the amendment to this initial decision in Decision 03/2022. This decision is often referred to as the platform *pricing methodologies*. The connection to the platforms, coupled with the pricing methodology, create an impetus for standardizing the balancing products and for transitioning to "pay-as-cleared" settlement (also known as marginal pricing), for whichever Member States this is not the case yet, in accordance with Article 30.1.a of the Electricity Balancing Guideline (EBGL). The push to harmonize balancing market operation to fully harness the benefits of the European integration also takes the form of the *imbalance settlement harmonization methodology* (ISHM) [ACER, 2020b]. This decision provides a framework for the calculation of the imbalance price. It prompts TSOs to adopt an imbalance pricing scheme based on the balancing prices generated by the platforms.

The sequence of integration, with balancing markets being the last, can be said to be in reverse of the appropriate order [Hogan, 2019]. Even though the volumes involved are much lower than the ones traded in day-ahead and intraday markets, the design that determines price formation in balancing and imbalance settlement should not be overlooked, as the expectation of the real-time prices generated by the balancing markets drives the wholesale electricity price in the day-ahead and other forward markets [Cervigni and Perekhodtsev,

[2] Italy has decided to suspend its participation in PICASSO starting on the 15th of March 2024.

2013]. As Hogan has stated repeatedly: "The last (price generated) should be (designed) first" [Hogan, 2016]. Additionally, the potential synergies brought by the integration of balancing are estimated to be considerably more significant than the ones from coupling day-ahead and intraday markets [Newbery et al., 2016]. The potential savings from *imbalance netting* and the use of cheaper balancing resources have been simulated to 400 million € for the Nordics alone [Farahmand and Doorman, 2012].

1.4 Scarcity Pricing through an Operating Reserve Demand Curve

In theory, short-run electricity markets should be able to support the efficient long-run investment that ensures the adequacy of the system. The short-run prices are set by the marginal cost of the marginal unit when demand meets supply and by the value of lost load (VOLL) when there is insufficient capacity and loads need to be curtailed. This forms a long-run equilibrium in which every technology breaks even [Boiteux, 1960]. A necessary component of this equilibrium is *scarcity pricing*, which refers to spot prices rising above the marginal cost of every unit present in the system in times of scarcity. This allows investors to recuperate their investment costs [Cramton and Stoft, 2005]. The equilibrium number of scarcity hours is directly connected to the *cost of new entry* (CONE), measure in € per MW per year, of peaking units and the VOLL. If a peaking unit, with marginal cost MC_P, is only profitable in periods of curtailment when scarcity pricing occurs, the adequate number of curtailment hours per year, H, is equal to the cost of new entry divided by the profit margin of the peaking unit, $VOLL - MC_P$.

$$H = \frac{CONE}{VOLL - MC_P} \tag{1.7}$$

Unfortunately, electricity markets suffer from a lack of responsive demand. Electricity loads cannot specify their reliability preferences and this hinders scarcity pricing. This market failure leads to *non-priced rationing* in which loads are not curtailed based on their willingness to pay but rather on emergency protocols. This also leads to the price in times of scarcity being set administratively and not by the willingness to pay of the last disconnected load. *Price caps* constraining electricity prices below the VOLL increase the number of curtailment hours required for investors to recuperate their investment cost, as shown in (1.7). *Missing money* arises when there is a mismatch between the number of hours curtailed and the scarcity prices. Inefficient long-term signals are not new [Joskow, 2008] but they have been exacerbated by the rapid large-scale integration of RES that are fuelled by government subsidies.

These concerns demonstrate the need for making the demand side of electricity systems more responsive. While adequacy concerns may be mitigated by

capacity remuneration mechanisms outside of the energy markets, such mechanisms may inhibit the adoption of demand response. If capacity markets are established without considering the potential rise in demand response, the short-term prices induced by the resulting capacity mix are likely to prevent investment in demand response infrastructure. Similar to responsive generation, investment in demand response processes and load management requires the stimulus of better scarcity pricing [Hogan, 2013]. Policy-makers should focus on the root cause of inefficient long-term signalling, inadequate scarcity prices, before trying to fix the symptomatic adequacy concerns through capacity mechanism [Hogan, 2019].

The introduction of an *operating reserve demand curve* (ORDC) is a mechanism for improving scarcity pricing and jump-starting demand response. This administrative demand curve introduces price elasticity in the procurement of reserve in the real-time market. Scarcity pricing induced by ORDCs exhibits less pronounced and more frequent price spikes than scarcity pricing induced by the *value of loss load* alone (VOLL pricing). The resulting "well-behaved" energy price creates a more stable environment for investment. Early references to this mechanism are found in Stoft in [Stoft, 2002] and the theory was formally anchored to the *loss of load probability* (LOLP) and the value of loss load by Hogan in [Hogan, 2005] and [Hogan, 2013]. Although scarcity pricing is a real-time mechanism, the uplifted real-time energy and reserve prices are expected to back-propagate to forward (e.g. day-ahead) energy and reserve markets, thereby generating robust investment signals in the market [Papavasiliou et al., 2021].

Scarcity pricing has been increasingly considered for implementation by a number of US independent system operators, such as ERCOT and PJM. These markets co-optimize energy and reserve and have been transitioning from an ORDC with a fixed reserve requirement to a downward sloping ORDC based on the VOLL and LOLP [NYISO, 2019]. ISO-NE and MISO are also considering this transition, following the recommendation of their respective market monitors. TSOs and regulators in Europe have also considered variations of scarcity pricing mechanisms. A *Reserve Scarcity Pricing* function based on the VOLL and the amount of leftover reserve based on hour-ahead measurements is implemented in Great Britain [Department for Business, 2020]. One objective behind the introduction of the measure was to generate scarcity prices without relying on agents internalizing inframarginal rents. The latter method for generating scarcity prices can be hard to distinguish from an exercise of market power. Ireland has also implemented a scarcity pricing mechanism, nevertheless it is not based on ORDC and LOLP but rather triggered by stress conditions in the system [SEM, 2021, EirGrid, 2017].

1.4.1 ORDC to Approximate a Stochastic Economic Dispatch

The introduction of an explicit ORDC to form an energy and reserve economic dispatch is motivated by the price generated by a two-stage stochastic economic dispatch as laid out by Hogan [Hogan, 2013]. His reasoning is replicated here. The objective of a central auctioneer is to maximize the first-stage welfare of serving the load plus the expected cost of the recourse action after the dispatch order. Let us assume an aggregate cost function, $C(\cdot)$, an aggregate marginal cost function, $MC(\cdot)$, an aggregated maximum technical production, Q, a load benefit function $B(\cdot)$, and a load marginal benefit function, $MB(\cdot)$. Let x be the net load change after the dispatch with probability density function $f(\cdot)$ and cumulative distribution function $F(\cdot)$.

$$\max_{q,d,r,s} \quad B(d) - C(q) + \int VOLL \cdot s(x) - \big(C(q + s(x)) - C(q)\big) dF(x) \tag{1.8a}$$

$$\begin{aligned}
s.t. \quad (\mu): \quad & q + r \leq Q && && (1.8b)\\
(\lambda): \quad & d = q && && (1.8c)\\
(\gamma(x)): \quad & s(x) \leq r && \forall x && (1.8d)\\
(\theta(x)): \quad & s(x) \leq x && \forall x && (1.8e)\\
& s, r \geq 0 && && (1.8f)
\end{aligned}$$

The stochastic cost relative to covering the net load change is composed of the benefit of serving the additional load valued at $VOLL$ and the additional production cost. The two-stage stochastic economic dispatch is presented in equation (1.8a) with q and d being the production and consumption in the dispatch respectively, r being an auxiliary variable introduced to represent reserve in the model, and $s(x)$ being the recourse additional production indexed by the net load change x. Note that only upward net load increases are considered in this analysis. The stochastic economic dispatch can be approximated by replacing the additional production cost with its first-order approximation, $\widehat{MC}(q) \cdot s(x)$.

$$\max_{q,d,r,s} \quad B(d) - C(q) + \int \big(VOLL \cdot s(x) - \widehat{MC}(q) \cdot s(x)\big) dF(x) \tag{1.9a}$$

$$\begin{aligned}
s.t. \quad (\mu): \quad & q + r \leq Q && && (1.9b)\\
(\lambda): \quad & d = q && && (1.9c)\\
(\gamma(x)): \quad & s(x) \leq r && \forall x && (1.9d)\\
(\theta(x)): \quad & s(x) \leq x && \forall x && (1.9e)
\end{aligned}$$

$$s, r \geq 0 \tag{1.9f}$$

The KKT conditions of this problem are presented in equations (1.10).

$$(d): \quad \lambda = MB(d) \tag{1.10a}$$

$$(q): \quad \lambda = MC(q) + \int \widehat{MC}'(q) \cdot s(x) dF(x) + \mu \tag{1.10b}$$

$$0 \leq r \perp \mu - \int \gamma(x) dF(x) \geq 0 \tag{1.10c}$$

$$0 \leq s(x) \perp \widehat{MC}(q) - VOLL + \theta(x) + \gamma(x) \geq 0 \tag{1.10d}$$

$$\tag{1.10e}$$

$\gamma(x)$ can be isolated to

$$\gamma(x) = \begin{cases} VOLL - \widehat{MC}(q) & \text{if } x \geq r \\ 0 & \text{else,} \end{cases} \tag{1.11}$$

and this allows us to rewrite λ, the dual variable relative to the first-stage market-clearing constraint (1.9c), as

$$\lambda = MC(q) + \int_{x \geq r} VOLL - \widehat{MC}(q) dF(x) \tag{1.12a}$$

$$= MC(q) + (1 - F(r)) \cdot (VOLL - \widehat{MC}(q)) \tag{1.12b}$$

$$= MC(q) + LOLP(r) \cdot (VOLL - \widehat{MC}(q)) \tag{1.12c}$$

by dropping the derivative of the marginal cost which would be negligible and by defining the loss of load probability, $LOLP(r)$, as the probability that the net load change exceeds the reserve available. Equation (1.12c) can be interpreted as the energy price being equal to the marginal cost of production plus a scarcity component. This scarcity component represents the cost associated with the net load exceeding the available reserve. If there is no risk of curtailment and the probability of x exceeding r is null, then the energy price reverts back to the marginal cost of production and there is no scarcity component. Introducing stochasticity in the economic dispatch allows a correction to the energy price by reflecting the risk of curtailment.

Scarcity components can be approximated with a deterministic economic dispatch by co-optimizing energy and reserve, as in equation (1.1), through the introduction an operating reserve demand curve. Let us define $BR(\cdot)$ as the benefit function of holding reserve and $MBR(\cdot)$ as the marginal benefit function of holding reserve. The deterministic reserve and energy economic dispatch can be formulated as follows.

$$\max_{q,d,r,dr} \quad B(d) + BR(r) - C(q) \tag{1.13a}$$

$$s.t. \quad (\mu): \quad q + r \leq Q \tag{1.13b}$$

$$(\lambda): \quad d = q \tag{1.13c}$$

$$(\lambda^R): \quad dr = r \tag{1.13d}$$

There are now two market clearing constraints: one for energy in equation (1.13c) and one for reserve in equation (1.13d). The analysis of the KKT conditions shows us that the reserve price is equal to the marginal benefit of reserve,

$$\lambda^R = MBR(r), \tag{1.14}$$

and that the energy price is equal to the marginal cost of production plus the marginal benefit of holding reserve,

$$\lambda = MC(q) + MBR(r). \tag{1.15}$$

The price of energy is thus coupled to the price of reserve. This is due to the maximum production constraint (1.13b) that links the dispatch of energy and the dispatch of reserve. This coupling ensures that an asset that is scheduled for both energy and reserve should be indifferent between offering either product. The consistency between the energy and reserve economic dispatch and the stochastic economic dispatch is ensured by setting the marginal benefit of reserve equal to the product of the loss of load probability and the value of lost load minus an estimation of the marginal cost of production, $\widehat{MC}$:

$$MBR(r) = LOLP(r) \cdot (VOLL - \widehat{MC}) \tag{1.16}$$

1.4.2 Scarcity Pricing without Co-Optimization of Energy and Reserve

In the absence of co-optimization of energy and reserves, an implicit ORDC can be used to compute *scarcity adders* based on the amount of reserve available in real time. This adder reflects the level of stress in the system, and would correspond to the price of reserve in an independent real-time market for reserve driven by the valuation of reserve demand. The coupling between reserve and energy prices would then be implemented by adding this adder as a price component to the real-time energy price. The introduction of a real-time market for reserve and the use of an adder on the energy price has been implemented in ERCOT [Hogan and Pope, 2020] and would be more realistic to implement in EU markets in the immediate future. The use of implicit adders and real-time markets for reserve can only reproduce co-optimization in a simple setting without ramping constraints.

1.4.3 Scarcity Pricing and Network Collapse

The procurement of operating reserve can be justified by the risk of total network collapse. Remember that the objective of balancing consumption and production is to prevent the collapse of the electricity network caused by frequency deviations. Frequency deviations lead to power plants tripping and disconnecting from the network. Total network collapses are fundamentally different from controlled rolling blackouts. When load is shed during controlled rolling blackouts, served loads value generation at the value of lost load. During a network collapse, generation has no value as no load is served. There is an externality that prevents all loads from being supplied even if there is generation available.

The risk of network collapse compels system operators to procure operating reserve as shown in [Joskow and Tirole, 2007]. The authors demonstrate the need for operating reserve through a two-stage model. In the first stage, capacity is procured before the realization of the load. In the second stage, the system operator decides how much of the load to cover and the ratio of capacity to keep available as reserve. The availability of the capacity is then revealed and there is a network collapse if the reserve level is insufficient to cover the generation unavailability. The potential unavailability and the risk of network collapse justifies investment in additional capacity and the procurement of operating reserve. The equilibrium is characterized by three modes at the second stage: off-peak, reserve curtailment and load shedding.

- In off-peak mode, all the load is covered and the capacity is not binding with regard to the level of reserve. There is excess capacity and sufficient reserve for preventing the network collapse. The marginal value of investment in capacity is equal to zero (or the commitment cost of reserve in Joskow and Tirole's model).

- In reserve curtailment mode, all the load is covered but the level of reserve is gradually reduced compared to the off-peak mode. Capacity is binding with regard to the level of reserve. The marginal value of investment in capacity is equal to its contribution to reduce the probability of network collapse.

- In load shedding mode, load is shed to ensure a minimum reserve requirement is met. At that level of reserve, the risk of network collapse is such that reserve is more beneficial to the system than covering load. The marginal value of investment in reserve apacity reserve in that mode is equal to the value of the load shed.

This analysis has shown that reserve is a public good. Generators and loads will consider the reliability provided by reserve asexogenous from their own policy. They have no incentive to procure reserve themselves and there is a need for a neutral system operator to procure it on their behalf (or to force them to procure it).

Additionally, the value of marginal investment in capacity described earlier can be interpreted as the implicit value for incremental investment in reserve. These values are very similar to the reserve prices generated by a co-optimization model. The price of reserve is very low in off-peak mode when the value of reserve on the ORDC is negligible. The price of reserve is equal to the VOLL in load shedding mode when the level of reserve is lower than the minimum reserve requirement.[3] The price of reserve is equal to the improved reliability it provides to the system in reserve curtailment mode.

There are however a couple of subtle differences between the two approaches. First, Joskow and Tirole rely on a reserve ratio relative to the load which can prove delicate to implement through an operating reserve market. The problem is particularly visible in the reserve curtailment mode where the system operator has to set administratively the exact reserve ratio that will prevent load shedding but will mobilize all the capacity. Any mistake when setting that ratio can result in price swings between zero (when the system operator sets the reserve ratio too low and there is an excess capacity) and the VOLL (when the system operator sets the reserve ratio too high and this results in load shedding). Reserve ratios result in inelastic demand curve which is at the root of this problem.

Secondly, the value of marginal investment in capacity proposed by Joskow and Tirole does not necessarily equate to the marginal value of reserve in the dispatch. The available capacity is not necessarily a good metric to estimate reliability in modern capacity mixes with high levels of RES.

1.5 Contribution and Structure of the Thesis

The thesis includes three independent works revolving around scarcity pricing and European balancing markets.

- Chapter 2 proposes a method for validating the calibration of ORDCs. The administrative parametrization of the demand curves is tested on a simulator of the Belgian short-term electricity market. The simulator accounts for the scheduling of power plants and for dispatching constraints, and has been validated against historical data. The objective of the analysis is to measure the tradeoff between the cost of operation and system reliability resulting from the ORDC calibration. Eight variants of ORDCs are considered, and this analysis serves as the basis for supporting a recommendation to the Belgian regulatory authority for the roll-out of scarcity pricing in Belgium [CREG, 2021].

- Chapter 3 is a bridge between scarcity pricing and balancing markets, as it discusses how scarcity pricing can be implemented in the European setting. It investigates the ambiguity stemming from the lack of

[3] Eq. (1.16) does not include a minimum reserve requirement but it is quite straightforward to include it.

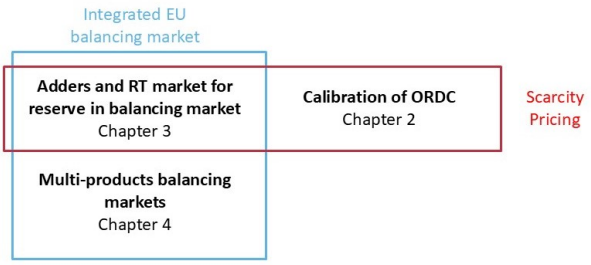


Figure 1.3: Thesis contributions.

co-optimization on the application of scarcity adders. Three variants are considered: (i) an adder on the imbalance price, (ii) an adder on the imbalance and balancing prices, and (iii) an adder on the imbalance and balancing prices in conjunction with a real-time market for reserve. The analysis demonstrates that adders without a real-time market for reserve induce inefficiencies by distorting the merit order and tend to be detrimental to the Member State that introduces them because the increased balancing cost is borne by the consumer in the zone with an adder.

- Chapter 4 provides a framework for analyzing the interaction of imbalance settlement with the clearing of multiple balancing energy and balancing capacity markets. The impact of (i) the imbalance settlement scheme, (ii) the mFRR activation strategy, and (iii) the capacity demand curves on the balancing market equilibria are assessed.

The works can also be categorized as a function of the methods that are used. Chapters 2 uses operations research techniques to optimize processes in a given policy setting. On the other hand, chapters 4 and 3 start at a lower level by discussing potential market designs. Game-theoretical techniques are used to predict the equilibrium effect of different market organizations.

The thesis contributions have been published in peer-reviewed academic journals, or are in the reviewing process.

- Chapter 2 is published in the *IEEE Transactions on Power Systems* under the title "Calibration of Operating Reserve Demand Curves Using a System Operation Simulator" [Cartuyvels and Papavasiliou, 2023]. An extension of the paper was published in the proceedings of the 2022 IEEE Power & Energy Society General Meeting (PESGM) under the title "Calibration of Operating Reserve Demand Curves using Monte Carlo Simulations" [**?**].

- Chapter 3 has been accepted for publication in the *IEEE Transactions on Energy Markets, Policy and Regulation* under the title "Market Equilibria in Cross-Border Balancing Platforms" [Cartuyvels et al., 2023].

- Chapter 4 is under review and a previous version can be found as a CORE discussion paper titled "Interactions of Imbalance Settlement with Energy and Reserve Markets in Multi-Product European Balancing Markets".

Additionally, I have participated as a second author in the publication of "Implementation of scarcity pricing without co-optimization in European energy-only balancing markets" in *Utilities Policy* [Papavasiliou et al., 2023] and as a first author in the publication of "Efficient Dispatch in Cross-Border Balancing Platforms: Elastic Demand Through Parametric Cost Function Approximation" as a proceeding of the 2024 International Conference on the European Energy Market (EEM) [Cartuyvels et al., 2024].

2 Calibration of ORDC Using a Short-Term System Operation Simulator

2.1 Introduction

The lack of a strong responsive demand side in electricity markets makes pricing in time of scarcity a delicate task that necessitates at least one administrative input to price electricity in times of scarcity [Cramton and Stoft, 2005]. *Energy-only* markets without explicit ORDCs need an estimate of the VOLL. Markets with explicit ORDCs also require the characterization of the loss of load probability. Additionally, markets relying on capacity markets to induce the efficient long-term capacity mix need to specify the capacity demand curves and participation guidelines. The list of administrative parameters for capacity markets includes penalty for not being present in times of scarcity and derating factors for renewable capacity.

The pre-implementation evaluation of these mechanisms' parametrizations is essential for policy makers. This chapter describes a short-term simulator for evaluating the calibration of ORDCs for the roll-out of scarcity pricing in Belgium. This tool is part of a number of studies performed on behalf of the Belgian regulatory authority for energy: [Papavasiliou and Smeers, 2017, Papavasiliou et al., 2018] assessed the potential of scarcity pricing based on ORDC, [Papavasiliou et al., 2021] used a stochastic equilibrium framework to test different levels of integration of scarcity in the Belgian context, and [Papavasiliou and Bertrand, 2021] investigated the use of adders to back-propagate balancing capacity prices in the Belgian balancing market.

The calibration of ORDCs has largely been restricted to *open-loop* analyses in the existing literature [Zarnikau et al., 2020, ELIA, 2018, Papavasiliou et al., 2018]. An open-loop analysis lacks feedback from the introduction of elastic ORDCs. It calculates adders based on a given a dispatch as opposed to a *closed-loop* analysis where dispatch would be affected by the ORDCs. The Belgian transmission system operator, ELIA, studies scarcity pricing in Belgium in [ELIA, 2018]. Papavasiliou et al. [Papavasiliou et al., 2018] provide a sensitivity analysis of the induced scarcity prices based on the parametrization of the

operating reserve demand curve. The scarcity prices are derived from the remaining reserve capacity of a unit commitment model that uses fixed reserve requirements. Zarnikau et al. [Zarnikau et al., 2020] analyze the impact of scaling ORDCs horizontally, as well as the effect of the ORDC on the real-time market price and investment incentives for natural-gas-fired generation in the Texas electricity market.

References [Zarnikau et al., 2020, Papavasiliou et al., 2018] highlight the influence of the shape of ORDC on the price and how the parametrization of the ORDC can affect the remuneration of different technologies through its effect on prices. Nevertheless, the open-loop approach proposed in those papers is not able to capture the dispatch and commitment incentives created by different calibrations of the ORDC. This shortcoming of open-loop analyses motivates our proposal for a nested modelling approach for simulating the short-term operation of the system and for modelling the interplay between reserve prices, operational efficiency, and uncertainty revelation.

Our simulator models the operation of a perfectly coordinated system, where a sequence of centralized optimization model commits and dispatches resources in a coordinated fashion. Uncertainty is assumed to stem from the actual load that needs to be served by the system. The sequential optimization of system scheduling aims at replicating the real-time controllability of the different assets present in the system, with a specific focus on quantifying the interplay between lags in decision making and the revelation of uncertain information in the system. The constraints on the scheduling of assets are inspired by the detailed commitment model of [Simoglou et al., 2010]. This allows us to quantify the fundamental tradeoff that ORDCs aim at balancing: incurring non-negligible fixed costs for committing flexible resources that can allow the system to operate reliably in real time, versus running the risk of not covering imbalances fully.

The proposed quantitative methodology could be used in the future to adapt the parametrization of ORDCs to the state of the system. An increase in renewable generation may introduce significant additional uncertainty to the system, resulting in wider ORDCs that can provide stronger signals for investment. Such a shift has already been discussed in ERCOT [Zarnikau et al., 2020] and our methodology can be used to assess such policies.

The nested modelling of system operation also appears in [Zhou and Botterud, 2014] and [Lavin et al., 2020], albeit less detailed. Nested models also appear in other contexts in [Simao et al., 2017], [Bakirtzis et al., 2015], [Daraeepour et al., 2019], and [Atakan et al., 2022]. We proceed to discuss the relation of these publications to our work.

Zhou and Botterud [Zhou and Botterud, 2014] develop their model in order to analyze an ORDC which is based on the loss of load probability, and which accounts for the uncertainty caused by wind and load forecast errors, as well as generation contingencies. Lavin et al. [Lavin et al., 2020] introduce an LOLP which is a function of the ambient temperature, in order to represent the higher probability of forced generator outages under extreme temperature

conditions. Those models use short-term system operation simulators in order to compare different ORDC schemes. Nevertheless, these publications mostly leave aside the aspect of calibrating the ORDC. More specifically, the results of Zhou and Botterud [Zhou and Botterud, 2014] and Lavin et al. [Lavin et al., 2020] are limited by the hourly temporal granularity of their model, as they cannot quantify the tradeoff between short-term dispatch adjustments and the lagged activation of reserves. The increased precision of our model remedies this shortcoming, and advances the state of the art by assessing more faithfully the interplay between costs and lags versus reliability of operation for different calibrations of ORDC in a setting with multiple reserve products.

In this sense, the precision of our model is closer to the Smart-ISO model of Simao and Powell [Simao et al., 2017]. The model of [Simao et al., 2017] is inspired by system operator practices and it is used in order to assess the reliability of the PJM system under different levels of wind power integration. We can also draw similarities with the model of Bakirtzis et al. in [Bakirtzis et al., 2015] where the authors propose a number of short-term operating models in order to cope with the increased uncertainties of power system operations. Daraeepour et al. [Daraeepour et al., 2019] have a model with multiple optimization problems to estimate the gain from adopting a stochastic model when committing assets. More recently, Atakan et al. [Atakan et al., 2022] use their model to compare the cost performance of different levels of reserve requirement. Their calibration of the level of reserve requirement based on simulations in a stochastic hierarchical planning framework is simular in spirit to our approach. Note, however, that none of the aforementioned papers is focused on the question of the calibration of ORDCs.

Discussions on the shape of demand curves also appear in the *capacity market* literature. Cramton summarizes lessons learned from the early capacity market designs, and provides recommendations for the design of capacity demand curves in [Cramton and Stoft, 2005]. References [Brown, 2018] and [Zhao et al., 2018] focus on the parametrization of capacity demand curves. Brown [Brown, 2018] derives curve parameters that aim at maximizing welfare in a 3-stage model where two firms can (i) choose their production capacity, (ii) compete in a capacity auction and (iii) compete in an energy market. Zhao, Zheng and Litvinov [Zhao et al., 2018] propose a framework for rigorously determining capacity demand curves based on the value of reliability obtained by simulating different levels of capacity in the market.

Our paper can also be linked to Mays [Mays, 2021]. The author advocates for a more direct connection between the construction of ORDCs and the expected outcome of a stochastic market.

The main contributions of this chapter are thus threefold. First, we provide a quantitatively sound methodology for calibrating ORDCs. Previous works in the scarcity pricing literature are either not accounting for the feedback effect of introducing ORDCs due to being open-loop [Zarnikau et al., 2020, Papavasiliou et al., 2018], or they are focusing on the comparison between fixed reserve requirements and price elastic ORDCs and not on the parametrization of such

ORDCs [Zhou and Botterud, 2014, Lavin et al., 2020]. Our method anchors the calibration to a closed-loop simulation model and examines three specific design criteria for ORDCs that emerge in a realistic implementation of the mechanism.

Second, we implement detailed system operation models for quantifying the tradeoff between incurring fixed costs for committing flexible resources and allowing the system to operate reliably instead of running the risk of shedding load.

Third, the current modeling effort is contributing directly to the implementation of a scarcity pricing mechanism in the Belgian electricity market. The results of the analysis constitute the basis for our recommendation to the Belgian regulatory authority for the possible implementation of scarcity pricing in Belgium in the short to medium term. This recommendation supports their latest study [CREG, 2021].

The chapter is structured as follows. Section 2.2 summarizes the principles of scarcity pricing, with a specific focus on certain dilemmas pertaining to the calibration of ORDCs. Section 2.3 presents the short-term operating model that we have developed in order to support the calibration of an ORDC. Our case study of the Belgian market is presented in section 2.4. The results of the case study are presented and analysed in section 2.5. Finally, we conclude and discuss future research perspectives in section 2.6.

2.2 Scarcity Pricing and ORDC

Scarcity pricing is summarized in 2.2.1. Subsection 2.2.2 considers the case of multiple reserve products and subsection 2.2.3 introduces variants of ORDC.

2.2.1 Rationale of Scarcity Pricing

Following 1.4.1, ORDCs can be used to replicate deterministically the dispatch and pricing from a stochastic economic dispatch [Hogan, 2013]. In this setting, the value of an additional MW of balancing capacity is linked to the value of the improved reliability that it provides to the system by reducing the likelihood of load shedding. This marginal value is characterised by Hogan in [Hogan, 2013] as a function of the value of lost load ($VOLL$), the loss of load probability ($LOLP(\cdot)$) given the level of reserve in the system (r) and the marginal cost of the marginal unit in the system ($\widehat{MC}$):

$$MBR(r) = (VOLL - \widehat{MC}) \cdot LOLP(r) \tag{2.1}$$

In general, the value of lost load represents the willingness to pay to avoid a power outage. Different consumers (residential, commercial and industrial) have different values of lost load depending on the electricity use-case and the electricity consumption context (outage duration, warning, ...) [Gorman, 2022].

The *VOLL* in this context is the average value of lost load for the agents that would be curtailed in a rolling blackout [Hogan, 2013].

The most complete integration of scarcity pricing based on ORDC to electricity market operations would correspond to the co-optimization of reserve and energy in real time. An ORDC based on (2.1) would then be the explicit demand curve for reserve and would be inserted in the multi-product auction that trades energy and reserve simultaneously. Co-optimized markets would then produce one price for each product: (i) a reserve price for the available reserve and (ii) an energy price for the energy traded. Note that, in the absence of binding ramp constraints, the price of energy will be coupled to the price of reserve in order to ensure an equivalence between the marginal profit on the energy and reserve market for a marginal generator that supplies both reserve and energy.

In the absence of a co-optimization of energy and reserves, we could use the expression of Eq. (2.1) to compute *adders* based on the amount of reserve that is available in real time, as measured by system telemetry. This adder reflects the level of stress in the system, and would correspond to the price of reserve. The coupling between reserve and energy prices would then be implemented by adding this adder as a price component to the real-time energy price in the absence of co-optimization (the balancing energy price in EU nomenclature).

2.2.2 Multiple Reserve Products

Formula (2.1) has been generalized by Hogan in [Hogan and Pope, 2019] to the case of multiple reserve products of different quality. The quality of reserve refers to the delivery time that is required for this specific reserve product to be fully available, which in EU jargon is referred to as *full activation time*. This generalization is based on the split of a real-time dispatch interval in two parts. In the first part of the interval, it is assumed that only high-quality resources can respond, whereas in the second part of the interval, all reserve types are assumed to be able to respond. We consider the full interval as an imbalance interval.

In the EU, the reference duration for an imbalance interval is 15 minutes and [Papavasiliou et al., 2019] suggests the following split of the real-time dispatch based on the products that have been historically available in the EU balancing market. The first part of the interval would last 7.5 minutes and imbalances would be resolved by balancing capacity that can be fully activated in no longer than 7.5 minutes (which corresponds in our analysis to aFRR capacity[1]). The second part of the interval would also last for 7.5 minutes. In this time interval, imbalances would be resolved by balancing capacity that can be fully activated in 15 minutes (which correspond in our analysis to mFRR capacity).

[1]Note that, even though the full activation time of aFRR that is envisioned in the pan-European platform PICASSO for the activation of aFRR capacity is 5 minutes, the current analysis is performed for the Belgian system where aFRR has been required to be fully activated in 7.5 minutes in the past [E-Bridge Consulting, 2014]

Based on this split of an imbalance interval, the authors in [Papavasiliou et al., 2019] suggest the introduction of two ORDCs: (i) a 7.5-minute ORDC (eq. (2.2)) for the first part of the interval and (ii) the 15-minute ORDC (eq. (2.3)) for the second part of the interval:

$$MBR_{7.5}(r_{7.5}) = \frac{1}{2} \cdot (VOLL - \widehat{\text{MC}}) \cdot LOLP_{7.5}(r_{7.5}) \tag{2.2}$$

$$MBR_{15}(r_{15}) = \frac{1}{2} \cdot (VOLL - \widehat{\text{MC}}) \cdot LOLP_{15}(r_{15}) \tag{2.3}$$

Here, $LOLP_x(\cdot)$ corresponds to the loss of load probability after x minutes, r_x is the amount of reserve that can be activated within x minutes, and $\widehat{\text{MC}}$ represents the marginal cost of the system. The loss of load probability after x minutes is described in equation (2.4) and represents the probability of the imbalance after x minutes exceeding the balancing capacity that can be made available in x minutes:

$$LOLP_x(r_x) = \mathbb{P}(imb_x \geq r_x) \text{ with } imb_x \sim \mathcal{N}(\mu_x, \sigma_x^2). \tag{2.4}$$

The ORDCs described in (2.2) and (2.3) are obtained by assuming that reserve that can be activated in 15 minutes can prevent load curtailment in the second interval (from 7.5 to 15 minutes) and that the marginal value of that reserve should represent the improved reliability it brings in that second interval. Reserve that can be activated in 7.5 minutes can prevent load shedding in both intervals so the marginal value of that reserve should represent the improved reliability it brings in both intervals. Note that the amount of reserve that can be activated within 15 minutes includes both the reserve that can be activated in 7.5 minutes and the reserve that can be activated in 7.5 to 15 minutes.

The imbalance is assumed to be drawn from a normal distribution with mean μ_x and standard deviation σ_x. These parameters can be estimated from the historical system imbalance. They are computed per 4-hour block and per season, in order to account for seasonality.

The settlement in a co-optimized market can be understood by analysing the following model in a convex setting. The model presented below is solely for the sake of explaining the settlement in a system with multiple products of reserve. Assuming a benefit function for demand ($B(\cdot)$) and a constant production cost for generator $g \in \mathcal{G}$ (C_g), where $\mathcal{G}$ is the set of generators in the system, our goal is to maximize the welfare of the system as a function of the demand (d), reserve available after 7.5 and 15 minutes ($r_{7.5}$ and r_{15})[2] and the production (q_g) and supply of fast and slow reserve (r_g^F and r_g^S) for every

[2]The value of r_{15} would be computed in practice ex post, based on telemetry measurements. In case the resolution of telemetry data is 15 minutes, it would be necessary to assume a pre-defined availability of different resources for 7.5 minutes, which is the case in our present study.

generator g:

$$\max_{\substack{d,p,r^S,\\ r^F,r_{7.5},r_{15}}} B(d) - \sum_g C_g \cdot q_g + \int_0^{r_{7.5}} MBR_{7.5}(x)dx + \int_0^{r_{15}} MBR_{15}(x)dx, \tag{2.5}$$

The co-optimization must obey the market clearing constraint for energy and fast and slow reserve (the associated dual variables are provided in parentheses):

$$(\lambda): \quad d = \sum_g q_g, \tag{2.6}$$

$$(\lambda^{7.5}): \quad r_{7.5} = \sum_g r_g^F, \tag{2.7}$$

$$(\lambda^{15}): \quad r_{15} = \sum_g (r_g^F + r_g^S), \tag{2.8}$$

The operating constraints of generator g are characterised by the set $\mathcal{X}_g$:

$$(q_g, r_g^S, r_g^F) \in \mathcal{X}_g. \tag{2.9}$$

It is worth noting that the fast reserve supplied by the generators is eligible not only for the pool of reserve available after 7.5 minutes (Eq. (2.7)) but also for the pool of reserve available after 15 minutes (Eq. (2.8)).

The profit maximization problem faced by a generator g can be obtained by first relaxing the market clearing constraints of the co-optimization model (Eqs. (2.6)-(2.8)):

$$\begin{aligned} \max_{\substack{d,q,r^S,\\ r^F,r_{7.5},r_{15}}} \; & B(d) - \sum_g C_g \cdot q_g + \int_0^{r_{7.5}} MBR_{7.5}(x)dx + \int_0^{r_{15}} MBR_{15}(x)dx \\ & + \lambda \cdot (\sum_g q_g - d) + \lambda^{7.5} \cdot (\sum_g r_g^F - r_{7.5}) \\ & + \lambda^{15} \cdot (\sum_g r_g^F + \sum_g r_g^S - r_{15}) \end{aligned} \tag{2.10}$$

$$(s.t.) \quad (q_g, r_g^S, r_g^F) \in \mathcal{X}_g, \quad \forall g \in \mathcal{G}, \tag{2.11}$$

and then decomposing the relaxed problem by $g \in G$:

$$\max_{q_g, r_g^S, r_g^F} \quad q_g \cdot (\lambda - C_g) + r_g^F \cdot (\lambda^{15} + \lambda^{7.5}) + r_g^S \cdot \lambda^{15} \tag{2.12}$$

$$(s.t) \quad (q_g, r_g^S, r_g^F) \in \mathcal{X}_g. \tag{2.13}$$

From the generator point of view, $\lambda, \lambda^{7.5}$ and λ^{15} are exogenous parameters representing respectively the price of energy, fast reserve, and slow reserve for the first and second interval of an imbalance period. From the system operator point of view, these prices are can be obtained by solving the initial co-optimization problem.

The prices of reserve for the first and second interval, $\lambda^{7.5}$ and λ^{15}, are set by the ORDCs. This is demonstrated by the KKT conditions of the initial co-optimization problem relative to the complementarity constraints implicating variables $r_{7.5}$ and r_{15} in (2.14) and (2.15):

$$\lambda^{7.5} - MBR_{7.5}(r_{7.5}) = 0 \tag{2.14}$$

$$\lambda^{15} - MBR_{15}(r_{15}) = 0 \tag{2.15}$$

Generators are then remunerated according to (2.12) under co-optimization.

- Balancing capacity that can be made available in 7.5 minutes is remunerated with the *fast reserve price*:

$$\lambda^F = MBR_{7.5}(r_{7.5}) + MBR_{15}(r_{15}). \tag{2.16}$$

- Balancing capacity that can be made available in 15 minutes is remunerated with the *slow reserve price*:

$$\lambda^S = MBR_{15}(r_{15}). \tag{2.17}$$

- Energy is remunerated in real time, with the energy price, λ.

A notable challenge of integrating scarcity pricing in the EU and certain past US markets is the lack of co-optimization in the real-time European market. Without co-optimization, a system operator cannot rely anymore on the dual variables of the co-optimization problem in order to characterize the prices and needs to resort to approximations. In the past, ERCOT has used adders as proxies to couple the reserve and energy prices, although the evolution of the ERCOT design is towards a co-optimization model [Hogan and Pope, 2020].

In the spirit of the original ERCOT design, [Papavasiliou et al., 2019] suggests to introduce 3 adders in order to approximate pricing under co-optimization. The *fast adder* and *slow adder* would be equal to the fast and slow reserve price and would remunerate balancing capacity that can be made available in 7.5 and 15 minutes. The *energy adder* would be equal to the fast adder and would be added to the real-time balancing price for remunerating energy.

Assuming that the energy adder is equal to the fast adder is an approximation that presumes non-binding ramp constraints. The reader is referred to the supplement of [Papavasiliou et al., 2021] for a more detailed discussion about this approximation. The explanation can be summarized as follows. The price of energy is fixed by the KKT conditions of the initial co-optimization problem relative to the generation variables q_g, r_g^F and r_g^S. Depending on which constraints of the model are binding, the energy price may be offset by a constant with respect to the fast adder, the slow adder, or neither.

2.2.3 Variants of ORDC

Eqs. (2.2) and (2.3) depend on a number of design parameters. Different variants of ORDCs can be produced depending on these assumptions. In this paper, we consider the following design parameters, which have been discussed in the context of the implementation of scarcity pricing in Belgium: (i) different values for VOLL, (ii) whether the argument of the LOLP operator is the reserve capacity remaining before or after the activation of reserve and (iii) whether imbalance increments within an imbalance interval are assumed to be correlated or not.

VOLL at 8300 €/MWh versus 13500 €/MWh

The Belgian federal planning bureau has estimated the Belgian VOLL at 8300 €/MWh in [Devogelaer, 2017]. This value has been used as the reference value of the VOLL in [Papavasiliou et al., 2019]. The value of 13500 €/MWh has also been suggested because it represents the current bidding limit of the imbalance price [CREG, 2018] and as such the market players' assumed highest VOLL.

Pre- versus Post-Activation

Reference [Papavasiliou et al., 2021] points out that the pre- and post-activation variants correspond to different interpretations of what making a certain quantity of reserve available in real time would mean in terms of system operator expectations. The pre-activation interpretation means that 1 MW of reserve implies that a resource has been afforded time to recover from its balancing dispatch during the previous imbalance interval. The post-activation interpretation means that the resource is prepared to offer 1 MW even if it has not been afforded time to return to its originally scheduled setpoint. The effect of the assumption is found to be significant in the context of the stochastic equilibrium formulation presented in [Papavasiliou et al., 2021]. As the time step of the real-time / balancing market becomes shorter (5 minutes currently in the US, and 15 minutes in Europe), the distinction becomes less relevant.

If the post-activation reserve capacity margin is denoted as r, then the pre-activation margin is $r - imb$, with imb being the difference between the scheduled and actual demand. This allows us to value balancing capacity at the beginning of an interval before absorbing the imbalance. In practice, these options correspond to different timings for computing the $LOLP$. The pre-activation variant computes the $LOLP$ at the beginning of the period whereas the post-activation variant does it at the end of the interval.

Independent versus Correlated Imbalance Increments

When splitting an imbalance period into two intervals, the full imbalance that needs to be covered also has to be split into two *imbalance increments*. Each interval is then responsible for causing one of the two imbalance increments.

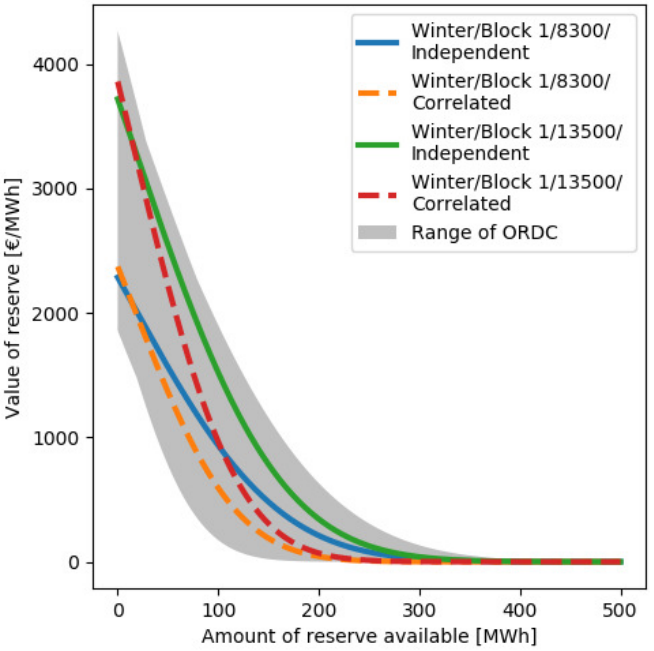


Figure 2.1: Example of ORDCs under different variants.

The assumption then is on how imbalance increments correlate in these two separate time steps. On the one extreme, we can assume correlated increments which implies that the total imbalance over both stages evolves linearly from the beginning to the end of the interval. On the other extreme, the assumption of independent imbalances implies that the total imbalance over both stages is the sum of two independently distributed imbalance increments occurring at stages 1 and 2 respectively.

The distinction affects the implied standard deviation of the imbalance that is used in the 7.5-minute version of Eq. (2.4). Given σ and μ, the standard deviation and mean of the 15-minute imbalance, the standard deviation of the 7.5-minute increments is either $\sigma/\sqrt{2}$ if the increments are independent, or $\sigma/2$ if the increments are perfectly correlated. The mean of both the independent and correlated 7.5-minute increments is $\mu/2$.

Fig. 2.1 presents different variants of ORDC. Each variant can be interpreted as a geometric transformation. The variation of VOLL, of the imbalance distribution, and of the activation time are respectively equivalent to vertical scaling, horizontal scaling and horizontal translation.

2.3 Simulator for Short-Term Operation

The structure of the simulator of short-term operation is described in subsection 2.3.1. We then characterize the generic unit commitment problem that is the building block of the different components of the simulator in subsection 2.3.2. The modifications to this generic unit commitment problem in order to account for the specificities of the intraday and real-time models are discussed in subsection 2.3.3.

2.3.1 Structure of the Simulators

The short-term operating model that we develop for the purpose of our analysis is composed of 4 embedded optimization problems that are solved in sequence throughout the day. Simulations begin in the day ahead by scheduling inelastic production, and unfold in intraday and real time by solving a sequence of unit commitment and economic dispatch problems with different scheduling windows. Each problem is employed for the commitment and / or dispatch of specific types of plants, depending on their response speed.

Particular care is given to (i) the operational constraints of the individually modelled generation plants, (ii) the revelation of real-time uncertainty and the scheduling of the system based on forecast information, and (iii) the effect of each decision-making stage on subsequent optimization problems.

Depending on the characteristics of an asset, its commitment plan and dispatch decisions will be obtained by different optimization problems. Assets can be partitioned into 3 broad categories, based on their real-time controllability.

1. **DA scheduled generators** cannot modify their planned day-ahead dispatch. This might be due to the inflexibility of the technology, or a link between electricity production and other processes, such as heating. The electricity production of these generators is typically determined in forward processes, and these units are not participating in a balancing market.

2. **Fast balancing capacity generators** require a non-negligible lag to start up (between 1 and 3 hours) but are very reactive once committed. CCGT generators constitute the bulk of this category. These assets can provide fast and slow reserve depending on their ramping constraint when committed.

3. **Slow balancing capacity generators** include all emergency generators and demand response resources. These generators are typically costly to start up, but can be activated in a short time, in order to free up some of the fast balancing capacity. These assets can provide fast and slow reserve when offline.

The 4 dispatch and commitment models that we develop optimize different subsets of the aforementioned assets. The dependencies are described hereunder. The sequencing of the models in the simulator is indicated in Fig. 2.2.

- **The day-ahead unit commitment (DA-UC) model** is used for scheduling the inelastic production that will not vary in real time relative to its day-ahead set-point. The model is launched once, before the beginning of the day, with a scheduling horizon of 72 hours. The model assumes a fixed initial dispatch of units for the day, which will be identical for every variant of ORDC that is tested in our analysis. The parameters of the simulation include the day-ahead load forecast, as well as settings that determine the reactivity and availability of the generation pool. This problem also determines the hydro storage target for the real-time models. The system is allowed to deviate from this target in order to address balancing issues, but such deviations are penalized.

- **The intermediate rolling-window unit commitment (Inter-RUC)** is solved every six hours over a 24-hour scheduling window. The Inter-RUC determines the commitment of CCGT plants for the next 6 hours until the next Inter-RUC is launched. This process thus proxies intraday market adjustments.

- **The pre-real-time rolling-window unit commitment (PRT-RUC)** determines the commitment of emergency generators. The model is launched every 15 minutes over a 1-hour scheduling window.

- **The real-time economic dispatch (RT-ED)** dispatches the generators that are committed in the previous optimization problems.

All of these models are necessary for simulating efficient system operation. The DA-UC allows us to account for uncertainty in day-ahead scheduling. The Inter-RUC provides real-time flexibility on the commitment of CCGT units. The PRT-RUC alone would be insufficient for those units because they require a long scheduling window to be worth committing. The PRT-RUC allows the system operator to react to the very short-term revelation of uncertainty by activating emergency measures. The RT-ED is necessary for determining a final position for the system, for quantifying scarcity adders, and for quantifying the cost of operating the system.

2.3.2 Generic Unit Commitment Problem

All the optimization problems are based on a standard unit commitment problem that aims at minimizing the total cost of the system under a series of constraints for both classical and pump-hydro generation.

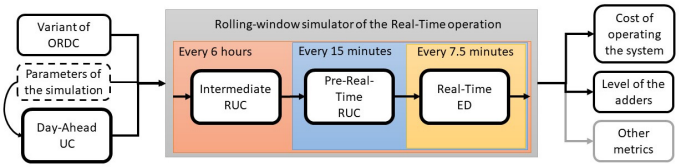


Figure 2.2: Sequence of models in our simulator of short-term operation.

Sets, Variables and Parameters

We define $\mathcal{T} = \{t_1, t_2, \dots t_T\}$ as the scheduling window for a problem, D as the demand, $\mathcal{G} = \{1, 2, \dots N\}$ as the set of generators and $\mathcal{S} = \{7.5, 15\}$ as the type of ORDC considered in this analysis. Let us also denote MBR_{it} as the marginal benefit function of reserve type i at period t. The segments of the ORDC are obtained by approximating (2.2) and (2.3) with a stepwise constant function[3].

The set of decisions concerning a generator g at period t is characterized by the point $x_{g,t} = (q_{g,t}, r^F_{g,t}, r^S_{g,t}, r^{NS,F}_{g,t}, r^{NS,S}_{g,t}, u_{g,t}, v_{g,t}, w_{g,t}, s_{g,t})$. This vector is the concatenation of the production, fast reserve, slow reserve, fast non-spinning reserve, slow non-spinning reserve and binary variables for the commitment, activation, shut-down and start-up of generator g at time t. The vector $x_{g,t}$ belongs to the set $X = \mathbb{R}^5_+ \times \mathbb{B}^4$.

Each generator g is characterized by its technical parameters P^+_g, P^-_g, R_g, $R^{NS,F}_g, R^{NS,S}_g, UT_g, DT_g$ and SU_g which are respectively the maximum and minimum production limit, the ramp rate, the limit for fast and slow non-spinning reserve, the minimum up time and down time, the start-up time of the unit, and its cost function $C_g : X \to \mathbb{R}$. The cost function includes the minimum load cost of keeping a generator online ($C^{ML}_g(\cdot)$), the start-up cost ($C^{SU}_g(\cdot)$), and the production cost with variable heat-rate ($C^P_g(\cdot)$). The cost function can be expressed as in Eq. (2.18):

$$C_g(q, u, v) = C^P_g(q) + C^{ML}_g(u) + C^{SU}_g(v) \tag{2.18}$$

The composition of the cost function is similar to the three-part bids of ERCOT with a production cost, a minimum load cost and a start-up cost.

1. The production cost is a function of production. It includes the emission cost, which depends on the emission price of a ton of CO_2 (C^{CO_2}) and the emission rate of the plant (ER_g). It also depends on the production cost which varies based on fuel price (C^F_g) and the variable heat rate as a function of the power output ($HR_g(p)$).

$$C^P_g(q) = \int_0^q C^{CO_2} \cdot ER_g + C^F_g \cdot HR_g(x)dx \tag{2.19}$$

[3] A method to approximate (2.1) by a stepwise constant function is proposed in [Zhou and Botterud, 2014].

2. The minimum load cost is a function of the commitment variable u and the minimum load of the plant (Q_g^{ML}) and the price of the fuel.

$$C_g^{ML}(u) = u \cdot Q_g^{ML} \cdot C_g^{Fu} \tag{2.20}$$

3. The start-up cost is a function of the activation variable v and the start-up load of the plant (Q_g^{SU}), as well as the price of the fuel.

$$C_g^{SU}(v) = v \cdot Q_g^{SU} \cdot C_g^{Fu} \tag{2.21}$$

The plant parameters are extrapolated from the real parameters of a subset of all the plants from a private database. This extrapolation is performed in order to adapt our realistic data to capacity levels that are not explicitly represented in our database. The fuel cost is based on the monthly historical spot TTF prices.[4]

We represent demand for energy and reserve using the vector $m_t = (z_t, r_t^{7.5}, r_t^{15})$. This tuple consists of the shortage in energy and the system supply for reserve for both the first and second half of an imbalance interval interval for period t. The hydro vector $h_t = (q_t^H, d_t^H, e_t^H, r_t^{H,F}, r_t^{H,S}, u_t^H)$ represents the production, consumption, energy stored, fast and slow reserve supplied by pumped hydro, and the pumping mode of a pumped hydro unit for period t. Note that $m_t \in \mathbb{R}_+^3$ and $h_t \in \mathbb{R}_+^5 \times \mathbb{B}^1$.

We further introduce the notation t_- to characterize the period preceding period t.

Objective Function

The system operator aims at minimizing the sum of the production cost and shortage cost minus the benefit from reserve:

$$\min_{\substack{x_{g,t},\\ m_t,h_t}} \sum_{t\in\mathcal{T}} \Big(\sum_{g\in\mathcal{G}} C_g(x_{g,t}) + VOLL \cdot z_t - \sum_{i\in\mathcal{S}} \int_0^{r_t^i} MBR_{it}(x)dx \Big). \tag{2.22}$$

The cost of shedding load is valued at the VOLL.

Market Clearing Constraints

$$D_t = \sum_{g\in\mathcal{G}} q_{g,t} + q_t^H - d_t^H + z_t \qquad \forall t \in \mathcal{T} \tag{2.23}$$

$$r_t^{7.5} = \sum_{g\in\mathcal{G}_D} r_{g,t}^F + r_{g,t}^{NS,F} + r_t^{H,F} \qquad \forall t \in \mathcal{T} \tag{2.24}$$

$$r_t^{15} = \sum_{g\in\mathcal{G}_D} r_{g,t}^S + r_{g,t}^{NS,S} + r_t^{H,S} + r_t^{7.5} \qquad \forall t \in \mathcal{T} \tag{2.25}$$

[4] Data available at `https://my.elexys.be/MarketInformation/SpotTtf.aspx`

Constraint (2.23) ensures that the market is balanced at all times, and constraints (2.24) and (2.25) define the market clearing conditions for fast and slow reserve. Reserve can be sourced from online generators (spinning reserve), from offline generators (non-spinning reserve) or from hydro (hydro reserve).

Generation Constraints

$$r^F_{g,t} \leq R_g \cdot 7.5 \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.26)$$

$$r^S_{g,t} \leq R_g \cdot 15 \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.27)$$

$$pq_{g,t} + r^F_{g,t} + r^S_{g,t} \leq P^+_g \cdot u_{g,t} \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.28)$$

$$p_{g,t} \geq P^-_g \cdot u_{g,t} \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.29)$$

Eqs. (2.26) to (2.29) represent the ramp constraints of fast and slow reserve and the maximum and minimum technical production limits of a unit.

Non-Spinning Reserve Constraints

$$r^{NS,F}_{g,t} \leq R^{NS,F}_g \cdot (1 - u_{g,t}) \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.30)$$

$$r^{NS,F}_{g,t} + r^{NS,S}_{g,t} \leq R^{NS,S}_g \cdot (1 - u_{g,t}) \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.31)$$

Eqs. (2.30) and (2.31) limit the supply of non-spinning reserve from offline generators. Most of the generators cannot provide non-spinning reserve and their parameters $R^{NS,F}_g$ and $R^{NS,S}_g$ are equal to 0.

Transition Constraints

$$q_{g,t} - q_{g,t\_} \leq R_g \cdot T \cdot (1 - v_{g,t}) + P^-_g \cdot v_{g,t} \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.32)$$

$$v_{g,t} + u_{g,t\_} - u_{g,t} - w_{g,t} = 0 \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.33)$$

The transition constraints (2.32) and (2.33) represent the ramp constraint for production and the commitment transition constraint. The production ramp constraint has two possible modes: one for normal operation and one for activation. The parameter T is the number of minutes of one period in a specific problem. For example, $T = 60$ for the DA-UC problem.

Operating Constraints

$$w_{g,t} + \sum_{t'=max(t_0, t-UT_g+1)}^{t} v_{g,t'} \leq 1 \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.34)$$

$$v_{g,t} + \sum_{t'=max(t_0,t-DT_g+1)}^{t} w_{g,t'} \leq 1 \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.35)$$

$$SU_g \cdot v_{g,t} - \sum_{t'=max(t_0,t-SU_g+1)}^{t-1} s_{g,t'} \leq 0 \qquad \forall g \in G, \forall t \in \mathcal{T} \qquad (2.36)$$

The operating constraints (2.34), (2.35) and (2.36) represent the minimum down time, minimum up time and start-up time of assets. Similarly as for T, UT, DT and SU are adapted to the granularity of the problem under consideration.

Hydro Generation Constraints

$$d_t^H \leq D_H^{Max} \cdot u_t^H \qquad \forall t \in \mathcal{T} \qquad (2.37)$$

$$e_t^H \leq E_H^{Max} \qquad \forall t \in \mathcal{T} \qquad (2.38)$$

$$q_t^H + r_t^{H,F} + r_t^{H,S} \leq P_H^{Max} \cdot (1 - u_t^H) \qquad \forall t \in \mathcal{T} \qquad (2.39)$$

$$q_t^H + r_t^{H,F} + r_t^{H,S} \leq e_t^H \qquad \forall t \in \mathcal{T} \qquad (2.40)$$

$$e_t^H = e_{t\_}^H - \frac{60}{T}(q_{t\_}^H + d_{t\_}^H \cdot \eta) \qquad \forall t \in \mathcal{T} \qquad (2.41)$$

The pumped hydro generation constraints restrict the maximum hydro consumption, energy stored and hydro production in constraints (2.37) - (2.39) with the pumped hydro characteristics D_H^{Max}, E_H^{Max} and P_H^{Max}. Note that a unit is either pumping or producing, as a function of the pumping mode u_t^H. Eq. (2.40) restricts the hydro reserve to the total stored energy. Constraint (2.41) describes the evolution of energy stored in the reservoir as a function of pumping and production decisions in the previous period, as well as the efficiency of the plant.

2.3.3 Modifications for the Intraday and Real-Time Problems

The intraday problem (Inter-RUC) and real-time problems (PRT-RUC and RT-ED) extend the standard unit commitment problem described previously to account for the day-ahead schedule and the particularities of the balancing market.

Day-ahead Constraints

$$r_{g,t}^F = 0 \qquad \forall g \in \mathcal{G}_I, \forall t \in \mathcal{T} \qquad (2.42)$$

$$r_{g,t}^S = 0 \qquad \forall g \in \mathcal{G}_I, \forall t \in \mathcal{T} \qquad (2.43)$$

$$q_{g,t} = q_{g,t}^{DA,*} \qquad \forall g \in \mathcal{G}_I, \forall t \in \mathcal{T} \tag{2.44}$$

Given the set $\mathcal{G}_I$ representing the DA-scheduled generators, constraints (2.42) and (2.43) restrict their ability to supply reserve. Constraint (2.44) characterizes their real-time inflexibility by equalizing the real-time production to the scheduled day-ahead production $p_{g,t}^{DA,*}$.

Status Constraints

The commitment decisions of previous optimization models are enforced by this set of constraints. We introduce the notation $\mathcal{S}_{U,t}$, $\mathcal{S}_{S,t}$, $\mathcal{S}_{A,t}$ and $\mathcal{S}_{F,t}$ to represent the set of generators that are unavailable, in the start-up process, activated or free in period t[5].

$$u_{g,t} = 0 \qquad \forall t \in \mathcal{T}, \forall g \in \mathcal{S}_{U,t} \cup \mathcal{S}_{S,t} \tag{2.45}$$

$$u_{g,t} = 1 \qquad \forall t \in \mathcal{T}, \forall g \in \mathcal{S}_{A,t} \tag{2.46}$$

$$s_{g,t} = 1 \qquad \forall t \in \mathcal{T}, \forall g \in \mathcal{S}_{S,t} \tag{2.47}$$

Constraints (2.45) and (2.46) imply that a generator is either off ($u_{g,t} = 0$) or on ($u_{g,t} = 1$) because of the minimum down time and minimum up time constraints of previous optimization problems. Similarly, Eq. (2.47) enforces the start-up variables dictated by a start-up decision in a previous problem and its start-up time.

Hydro-Deviation Constraints

The opportunity cost of hydro in real time is modelled by a hydro storage target $e_t^{DA,*}$ and the variable l_t representing the deviation from that target:

$$l_t \geq e_t^{DA,*} - e_t^H \qquad \forall t \in \mathcal{T} \tag{2.48}$$

$$l_t \geq e_t^H - e_t^{DA,*} \qquad \forall t \in \mathcal{T} \tag{2.49}$$

$$l_t \geq 0 \qquad \forall t \in \mathcal{T} \tag{2.50}$$

The term $\sum_{t \in \mathcal{T}} \int_0^{q_t} HD(q)dq$ is subtracted from the objective function of the real-time problem in order to penalize deviations.

Start-Up Constraints

The scheduling window for the pre-real-time problem is defined as $\mathcal{T} = \{t_{0,0}, t_{0,1}, t_1, \ldots, t_{w-1}\}$. It represents two 7.5-minute periods ($t_{0,0}$ and $t_{0,1}$) and $w - 1$ 15-minute periods. This window plans over $w \cdots 15$ minutes. The first period is split, in order to account for the start-up profile of emergency generators and how much of their generation is available after 7.5 minutes. Given the initial

[5] No generator can be in two sets at the same time and all generators must be in a set at every period t.

position of a generator p_g^0, the initial transition constraint of equation (2.32) needs to be reformulated:

$$p_{g,t_{0,0}} - p_g^0 \leq R_g \cdot 7.5 \cdot (1 - v_{g,t}) + R_g^{SU,0} \cdot v_{g,t} \quad \forall g \in G, \forall t \in \mathcal{T} \tag{2.51}$$

$$p_{g,t_{0,1}} - p_{g,0,1} \leq R_g \cdot 7.5 \cdot (1 - v_{g,t}) + R_g^{SU,1} \cdot v_{g,t} \quad \forall g \in G, \forall t \in \mathcal{T} \tag{2.52}$$

$$u_{g,t_{0,0}} = u_{g,t_{0,1}} \quad \forall g \in G, \forall t \in \mathcal{T} \tag{2.53}$$

$$v_{g,t_{0,0}} = v_{g,t_{0,1}} \quad \forall g \in G, \forall t \in \mathcal{T} \tag{2.54}$$

The start-up constraints (2.51) and (2.52) ensure that generators comply with their start-up profile. This start-up profile is characterized by the maximum production 7.5 minutes and 15 minutes after activation ($R_g^{SU,0}$ and $R_g^{SU,1}$). The start-up ramp profile is similar to the limit on non-spinning reserve for emergency generators and demand response.

Note that the two 7.5-minute dispatch periods only account for one 15-minute commitment period (Eqs. (2.53)-(2.54)).

Economic Dispatch Constraints

The real-time economic dispatch is similar to the first and second period of the pre-real-time unit commitment except that all the generators are either activated, unavailable or in start-up. No commitment decision is taken in this problem.

2.4 Model Validation against Historical Data

Subsection 2.4.1 describes the case study of the simulation and subsection 2.4.2 the validation of the day-ahead module.

2.4.1 Case Study

The investigation is performed on the Belgian power system with the historical load of 2018.

Generation pool

The generation pool modelled in the simulator includes all the controllable assets of Belgium and the load flexibility considered by the Belgian transmission system operator, Elia, and is presented in Table 2.1.

The controllable assets are mainly based on the database of installed capacity by unit, which is publicly available on the Elia open data platform. In

Table 2.1: Generation pool of Belgium

Balancing capacity category	Type of plant	Number of units	Max. agg. prod. [MW]
DA sched.	Cogeneration, run of river hydro, waste incinerators, Nuclear		7500
Fast bal. cap.	CCGT	10	3754
Slow bal. cap.	OCGT	6	302
Slow bal. cap.	Turbo-jet	10	194
Slow bal. cap.	Demand response		250 to 500
Slow bal. cap.	Foreign bal. cap.		50
Fast bal. cap.	Pump-hydro		1300

addition to these resources, this pool also includes demand response, pump-hydro and some foreign balancing capacity.[6] Demand response varies from month to month and is extracted from historical data. Foreign balancing capacity represents the availability on the transmission lines between ELIA and its neighbouring TSOs. This value is fixed at 50 MW following [Papavasiliou et al., 2019].

The technical parameters relative to the operating constraints of the flexible generators are largely aligned with [ELIA, 2019], except for the minimum production of emergency generators. We consider emergency generators and demand response as "all-or-nothing" generators.

Net Load

Net load is modelled as the power that must be served by flexible and controllable assets. It corresponds to the difference between grid load and renewable energy and imports / exports.

The data that we use for net load is obtained from ELIA's website and the ENTSO-E transparency platform.[7] The data resolution of the ELIA website and ENTSO-E platform is respectively 15-minute and hourly.

[6] Demand response and foreign balancing capacity are referred to as *Non-CIPU generation* and *Inter-TSO* in the Belgian framework.

[7] Data available at `https://www.elia.be/fr/donnees-de-reseau/data-download-page` and `https://transparency.entsoe.eu/`

Imbalance

The mean and standard deviation of the distribution of the imbalances used for the LOLP in (2.4) is obtained from the historical system imbalance recorded in [ELIA, 2021a]. For simulating 2018, we use the historical imbalance of the 3 preceding years. The full characterization of the mean and standard deviation of the system imbalance can be found in the appendix.

2.4.2 Validation

We validate our model by assessing the quality of the forward position computed by the day-ahead unit commitment compared to the historical records of day-ahead positions. The validation is restricted to the day-ahead unit commitment because the co-optimization of reserve and energy in real time in our simulator is a closed-loop investigation that is expected to produce a different dispatch depending on the ORDC that we analyse.

The comparison is performed over the aggregated forecast production per type of fuel. There are five types of fuel, namely (i) nuclear, (ii) gas, (iii) hydro, (iv) liquid fuel and (v) other. We will mostly focus on gas and hydro production. Nuclear and other technologies are mainly driven by the maximum available output and liquid fuel is used as an emergency measure and is rarely scheduled in the day ahead.

The performance of the simulator is compared to that of [Papavasiliou and Smeers, 2017]. Whereas [Papavasiliou and Smeers, 2017] aims at computing scarcity prices and does not perform a sensitivity analysis on the shape of the ORDC, both the present model and that of [Papavasiliou and Smeers, 2017] are calibrated against the Belgian system. The current work improves [Papavasiliou and Smeers, 2017] by (i) refining and extending the generation pool, (ii) reducing the granularity of the dispatch, and (iii) proposing a more realistic modeling of the dispatch and commitment decisions. These enhancements allow us to analyse the tradeoff between the commitment of fast balancing capacity and the cost of operating the system with more realism. Table 2.2 details this comparison and demonstrates that the increased modeling detail does not come at the cost of accuracy in replicating past observations of the Belgian electricity system.

Discrepancies between our simulations and the historical dispatch may originate from: (i) The difficulty of assessing the opportunity cost of pump-hydro and thermal assets that belong to a portfolio, (ii) the impact of climatic conditions on the availability of certain assets and (iii) difference in the commitment of particular CCGTs between the simulated and historical dispatch.

2.5 Results and Analysis

The results presented in this section are obtained by simulating the historical demand of Belgium for 2018. The analysis is based on a reference scenario and

Table 2.2: Mean Error (ME), Mean Absolute Error (MAE) and Root Mean Squared Error (RSME) between the historical and simulated production per type of fuel for 2018 and comparison with the errors of the study in [Papavasiliou and Smeers, 2017] for 2013.

		Gas	Hydro	Fuel	Nucl.	Other
ME	Simulator	-76.7	28.5	0.0	2.7	140.4
	1st Study	168.9	4.7			
MAE	Simulator	208.8	69.2	0.0	36.8	148.7
	1st Study	240.7	61.6			
RMSE	Simulator	267.7	113.7	1.0	116.2	176.9
	1st Study	309.9	119.3			

on a sensitivity analysis on the availability of the slow balancing capacity for contributing towards satisfying the demand of the 7.5-minute ORDC.

Our analysis focuses on comparing the total operating cost of the different variants of ORDCs (subsection 2.5.1), and on analyzing the impact of these variants on the level of the scarcity adder (subsection 2.5.2). The comparison focuses largely on the influence of the level of conservatism of the variants. More conservative variants (value of lost load at 13500 €/MWh and / or independent 7.5-minute imbalance increments) are compared against less conservative variants (value of lost load at 8300 €/MWh and / or correlated 7.5-minute imbalance increments).

Some of the slow balancing capacity is assumed to partly cover the demand of the 7.5 minute ORDC and [ELIA, 2018] highlights the importance of this assumption. We investigate further in this direction with the sensitivity analysis that is performed in section 2.5.3. Note that the reference scenario assumes an availability of 28%.

2.5.1 Cost Analysis of the Reference Scenario

The total cost of the variants is reported in Table 2.3. The values reported here are obtained by adding the production cost, minimum load cost, activation cost and shortage cost of the system, and do not include the cost of the price-inelastic generators, since the latter is identical across different scenarios. The last columns report the cost of emergency measures. The total cost varies from 1.697 M € per day to 1.683 M € per day. Thus, we find a difference of up to 14 k € per day between the different variants. This corresponds to a variation of up to 0.8% of the mean total flexible cost, which can be considered quite stable.

Despite the stability of the total cost, we can analyse the differences between the variants in order to better understand the impact of the ORDC on the com-

Table 2.3: Decomposition of the mean total operating cost of each variant in million € per day.

			Total cost	Prod. cost	Min. load cost	Act. cost	Short. cost	Em. cost
8300	Pre-Act.	Ind.	1.694	1.317	0.344	0.032	0.001	0.109
		Corr.	1.697	1.321	0.342	0.033	0.000	0.114
	Post-Act.	Ind.	1.691	1.316	0.343	0.032	0.000	0.107
		Corr.	1.694	1.318	0.343	0.033	0.000	0.110
13500	Pre-Act.	Ind.	1.688	1.304	0.352	0.031	0.000	0.101
		Corr.	1.687	1.302	0.350	0.033	0.002	0.097
	Post-Act.	Ind.	1.684	1.301	0.352	0.031	0.000	0.097
		Corr.	1.683	1.299	0.350	0.033	0.000	0.095

mitment and dispatch decisions. Based on Table 2.3, we can observe that more conservative variants are typically less costly. This trend is more accentuated for the variation of the VOLL, where the 13500 variants are consistently lower in cost than their 8300 counterpart. More conservative ORDCs tend to result in higher minimum load costs, and this is balanced out by their lower production cost. This is particularly the case when comparing the variants with different value of lost load. For instance, the 13500/Post-activation/Independent variant incurs 352,000 € of minimum load cost and 1,304,000 € of production cost, compared to the 8300/Post-activation/Independent that incurs 343,000 € of minimum load cost cost and 1,316,000 € of production cost.

2.5.2 Price Analysis of the Reference Scenario

The values of the adders that result from the different variants under the reference scenario are presented in Table 2.4. The fast adder varies from 2.7 €/MWh to 6.5 €/MWh and the slow adder from 0.15 €/MWh to 0.5 € per MWh. The adders generated by the different variants are thus more significantly dependent on the choice of ORDC than system cost. Two main observations can be highlighted from the table.

Firstly, conservative ORDCs (13500 variants and Independent variants) produce higher adders than their counterparts. Note that the most significant difference is caused by the distribution of the 7.5-minute imbalance increments, with the independent variants producing fast reserve adders that are approximately twice the value of their counterparts.

Secondly, correlated variants result in a higher slow reserve adder. This is driven by the fact that CCGTs have lower incentives for commitment, which

Table 2.4: Mean level of the adders for the reference scenario in €/MWh.

			Fast reserve adder	Slow reserve adder
8300	Pre-Act.	Ind.	5.78	0.25
		Corr.	2.86	0.36
	Post-Act.	Ind.	5.78	0.14
		Corr.	2.74	0.30
13500	Pre-Act.	Ind.	6.50	0.37
		Corr.	3.28	0.56
	Post-Act.	Ind.	6.20	0.21
		Corr.	2.92	0.32

decreases the committed balancing capacity and increases the value of the slow adder.

This last point highlights a fundamental difference between the variations in terms of distribution of imbalance increments versus the variations of the VOLL. The independent and correlated variants only impact the 7.5-minute ORDC and increase or decrease the incentives for committing CCGTs, while keeping the slow reserve demand constant. In comparison, variations of the VOLL impact both the 7.5-minute and 15-minute ORDCs.

Fig. 2.3 provides an indication about the persistence of the price signal generated by the ORDC in terms of profitability for owners of flexible assets. The figure compares the price signal obtained by 4 variants, beginning with the most conservative variant that produces the highest adder (13500/Post-activation/Independent) and modifying each of the design parameters in turn. The y-axis displays the measure of the adder under the metric of conditional value at risk as a function of the risk aversion of the agents on the x-axis. Depending on the risk aversion α of an agent, the agent will only consider the $100 - \alpha$ worst adders for computing its expected payoffs from the adder. The risk aversion can range from 0% to 100%, where 0% is a completely risk-neutral agent and 100% is a completely risk-averse one.

We observe a notable drop in the value of the payoff curve for low values of the x axis, which corresponds to the impact of a very high adder resulting from very stressed conditions in the system. These highly stressed conditions constitute less than 1% of the total possible outcomes in the system. It is possible to assess the quality of the signal produced by a variant by analysing the persistence of the adder when the risk aversion increases.

In Fig. 2.3 we observe that the correlated variant is the least persistent by a wide margin. The variants related to the value of lost load and the pre/post-activation capacity produce similar levels of persistence. Note that, for these variants, the decrease can be considered constant until a risk aversion level of

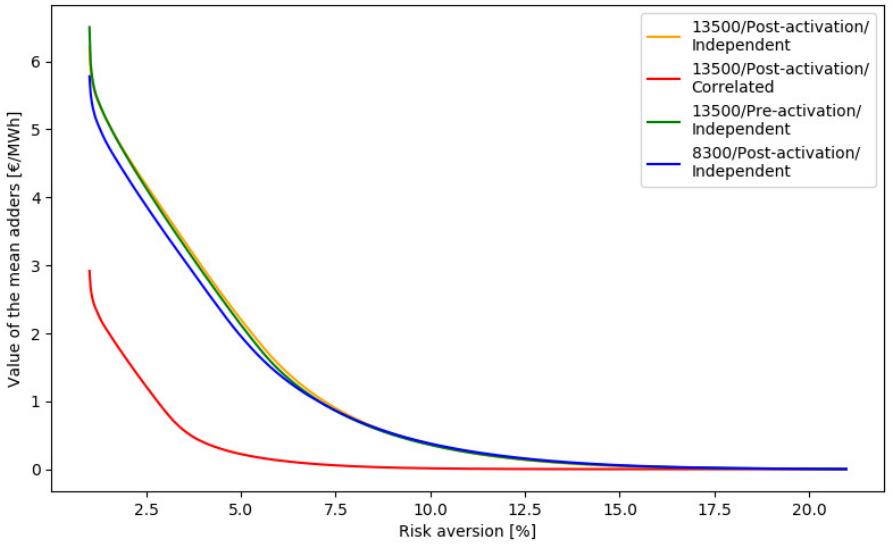
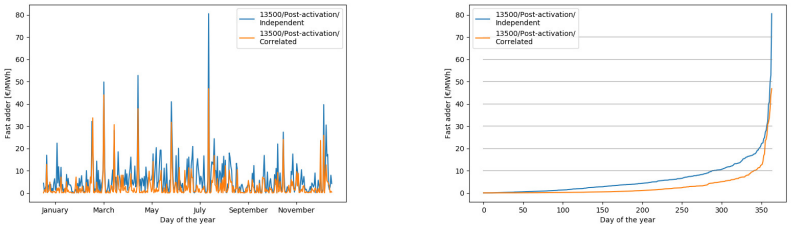


Figure 2.3: Adder payoff as a function of the risk aversion of the agents.



(a) (b)

Figure 2.4: (a) Yearly distribution and (b) cumulative distribution of the mean fast reserve adder per day [€/MWh] for 2018.

7.5%, which indicates a mean adder that is generated by the repetition of a large number of occurrences of small adders in the market, which is desirable from the perspective of mitigating investment risk.

Fig. 2.4 presents the mean fast adder per day for 2018. It shows that the mean price per day is between 0 and 10 €/MWh for the majority of the time, while during approximately 50 days the average adder is higher than 10 €/MWh. Scarcity adders turn out to be zero between 75 and 86% of the time in the simulations depending on the variant.

Table 2.5: Fast and slow reserve adder in €/MWh as a function of ρ, the availability of the slow balancing capacity for covering the demand of the 7.5-minute ORDC.

			Fast reserve adder			Slow reserve adder		
	ρ		0%	28%	50%	0%	28%	50%
8300	Pre-Act.	Ind.	14.65	5.78	1.57	0.12	0.25	0.33
		Corr.	12.88	2.86	0.81	0.19	0.36	0.54
	Post-Act.	Ind.	14.62	5.78	1.51	0.08	0.14	0.29
		Corr.	13.33	2.74	0.74	0.14	0.30	0.48
13500	Pre-Act.	Ind.	14.76	6.50	1.66	0.25	0.37	0.32
		Corr.	12.92	3.28	0.90	0.27	0.56	0.62
	Post-Act.	Ind.	14.91	6.20	1.55	0.09	0.21	0.25
		Corr.	13.12	2.92	0.92	0.17	0.32	0.62

2.5.3 Sensitivity Analysis for the Variation of the Availability of Slow Balancing Capacity for the 7.5-Minute ORDC

The eligibility of slow balancing capacity for the 7.5-minute ORDC is an important determinant for the remuneration of flexible assets. In principle, the eligibility of these resources should be plant-dependent and should reflect as accurately as possible their reactivity. For reasons of simplicity, the current Belgian scarcity pricing proposal only assumes a generic value for the availability. ELIA [ELIA, 2018] assumes an availability of 50%, and our analysis considers availabilities ranging from 0% to 28%.

Given an availability ρ, the parameters $R_g^{SU,0}$ in Eq. (2.51) and $R_g^{NS,F}$ in Eq. (2.30) for the emergency generators and demand response are modified as follows:

$$R_g^{SU,0} = R_g^{NS,F} = \rho \cdot R_g^{NS,S}. \tag{2.55}$$

The effects of modifying the availability on the adder are two-fold, and can be observed in Table 2.5. Increasing ρ (i) reduces notably the level of the fast adder by increasing the fast balancing capacity pool, and (ii) increases marginally the level of the slow reserve adder. Increasing the availability of mFRR for covering the demand of the 7.5-minute ORDC reduces the need for aFRR from CCGTs, and has a direct effect on their commitment. This compresses the committed balancing capacity, which in turn increases the level of the slow adder.

2.6 Conclusion

We develop a detailed unit commitment and economic dispatch simulation model of the Belgian power system in order to analyze the effect of different design choices for Operating Reserve Demand Curves on the cost of system operation and the price of aFRR and mFRR capacity. Our simulator attempts to emulate a best-case, fully coordinated operation of the system from the day ahead to real time. We propose four modules that are interleaved and implemented as a rolling horizon optimization.

The precision of our model allows us to account for the tension between incurring fixed costs for committing flexible resources that can allow the system to operate reliably in real time, versus running the risk of shedding load.

We validate our model against historically observed data of the Belgian market for 2018. We then perform a case study on the impact of ORDCs on scarcity prices and system costs for 2018. We also perform a sensitivity analysis on the extent to which mFRR reserves are assumed to contribute towards satisfying the demand for 7.5-minute reserves.

The main findings of our analysis can be summarized as follows:

1. The total flexible operating cost for a day is stable, regardless of the chosen ORDC variant. It is also stable for the specific generation pool of Belgium that is investigated in our work.

2. The fast adder varies from 2.8 €/MWh to 6.5 €/MWh in the reference scenario. The main driver of the price is the assumption related to the distribution of the 7.5-minute imbalance increments, followed by the value of lost load.

3. The level of the fast adder is sensitive to assumptions about what resources can contribute towards covering the demand of the 7.5-minute ORDC. Note that this sensitivity was already reported in [ELIA, 2018].

Note that, even though the methodology is generic, the conclusions regarding shape of the ORDC are system-dependent and cannot be generalized to other systems.

In future work, we are interested in developing a Monte Carlo simulation model for the Belgian system which draws samples of system uncertainty, instead of relying on historical data. This would allow us to enhance the statistical reliability of our results, by exposing the system to multiple years of hypothetical operation.

Table 2.6: Mean and standard deviation of the 15-minute imbalance distribution.

Season	Time of the day	Mean	Standard deviation
Winter	Block 1	29.00	160.25
Winter	Block 2	25.93	134.12
Winter	Block 3	6.77	165.30
Winter	Block 4	44.00	190.88
Winter	Block 5	56.95	169.15
Winter	Block 6	3.99	144.29
Spring	Block 1	7.74	145.75
Spring	Block 2	27.05	128.75
Spring	Block 3	-0.86	143.95
Spring	Block 4	28.81	173.13
Spring	Block 5	40.64	159.02
Spring	Block 6	-7.44	127.18
Summer	Block 1	14.54	134.15
Summer	Block 2	27.89	111.75
Summer	Block 3	0.86	130.06
Summer	Block 4	28.98	151.59
Summer	Block 5	27.60	144.17
Summer	Block 6	-5.93	119.16
Autumn	Block 1	11.62	151.34
Autumn	Block 2	29.19	124.09
Autumn	Block 3	-21.08	160.09
Autumn	Block 4	-7.58	175.77
Autumn	Block 5	-5.30	144.98
Autumn	Block 6	-10.95	150.09

2.A Seasonal Imbalance

The parameters of the ORDCs, which are extracted from the historical seasonal imbalances, can be found in Table 2.6. Block 1 is 10 pm to 2 am, block 2 is 2 am to 6 am, block 3 is 6 am to 10 am, block 4 is 10 am to 2 pm, block 5 is 2 pm to 6 pm and block 6 is 6 pm to 10 pm.

3 Adders and Market for Reserve in Integrated Balancing Markets

3.1 Introduction

The *imbalance settlement harmonization methodology* (ISHM) offers to the TSOs the possibility of unilaterally introducing a *scarcity component* or an *incentivizing component* in their imbalance price. Such adders on the imbalance price are already present in Belgium and the Netherlands, with the stated goal of inducing flexibility consumers to contribute towards balancing the system. They aim at incentivizing flexibility consumers to participate in the balancing process by penalizing (resp. rewarding) positions that hurt (resp. help) the system. Imbalances in the same (resp. opposite) direction as the system imbalance are punished (resp. rewarded) with an increased imbalance price. Similarly the possibility of introducing a scarcity component to the balancing prices generated by the platforms is foreseen by the cross-border balancing platforms' *pricing methodology*. Adders on the imbalance or balancing prices do not target the same agents. Adders on the imbalance price are paid by flexibility consumers or *balancing responsible parties* (BRPs) and adders on the balancing price are paid to the flexibility providers or *balancing service providers* (BSPs).

The intention of introducing adders to the imbalance price is to keep the system imbalance stable and to prevent long-lasting imbalances by BRPs [ELIA, 2021b]. This imbalance pricing scheme holds BRPs accountable for their consumption of balancing capacity and one possible interpretation is that this acknowledges the real-time value of balancing capacity. The introduction of adders on the imbalance and/or the balancing price can thus be interpreted as an incomplete implementation of scarcity pricing through an operating reserve demand curve. The lack of real-time markets for balancing capacity in European balancing markets generates ambiguity on the application of scarcity components and this chapter investigates the direct distributional effect triggered by these design choices.

This work uses an analytical model to investigate how adders foreseen by the ISHM and the pricing methodology can be applied in European balancing markets connected with cross-border balancing platforms. Three designs

are examined: (i) The *adder on BRPs* design, which is currently used in Belgium by ELIA, the Belgian TSO, where the adder is applied on the imbalance price [ELIA, 2022], (ii) the *adder on BRPs and BSPs* design, suggested by the Dutch TSO TenneT, where an adder is applied on both the balancing and imbalance price [Nederlands, 2023], and (iii) the *Real-Time (RT) market for reserve* suggested by [Papavasiliou, 2020, Papavasiliou et al., 2021, Papavasiliou and Bertrand, 2021]. This design proposes to remunerate available but non-activated balancing capacity in addition to balancing energy. The coupling of the balancing capacity and balancing energy market uplifts the balancing and imbalance price by an adder equal to the balancing capacity price. The value of balancing capacity, the *reserve price*, is based on an ORDC which represents the probability of losing load given the current state of the system [Hogan, 2005].

This analysis is part of the balancing market design literature. Seminal work on coupled capacity and energy auctions includes [Bushnell and Oren, 1994], where the authors identify with an analytical model the optimal bidding strategy and the necessary conditions for an equilibrium in early bid scoring systems with discriminatory settlement rules. Chao and Wilson establish, by arguing through backward induction, that uniform pricing for both energy and reserve can create incentives for truthful bidding [Chao and Wilson, 2002]. Similar analytical methods have been employed in order to investigate the interplay between the wholesale market and the balancing market [Ehrhart and Ocker, 2021] and the switch from "pay-as-bid" to "pay-as-cleared" auctions [Ocker et al., 2018]. Agent-based [van der Veen et al., 2012, Poplavskaya et al., 2020, Poplavskaya et al., 2021, Papavasiliou and Bertrand, 2021] and simulation-based methods [Petitet et al., 2019] have also been used for analysing the impact of the imbalance pricing scheme on system cost [van der Veen et al., 2012], strategic bidding in joint or split balancing capacity and balancing energy markets [Poplavskaya et al., 2020, Poplavskaya et al., 2021], the back-propagation of real-time prices to day-ahead prices [Papavasiliou and Bertrand, 2021], and the timing of the gate closure time [Petitet et al., 2019].

There is also a broad literature on optimizing the strategy of different agents in balancing markets. Such literature includes analyses of optimal trading strategies [Matsumoto et al., 2021, Bottieau et al., 2021], the minimization of portfolio imbalance [Hellmers et al., 2016], the optimal activation of balancing energy by system operators [Shinde et al., 2021], and the optimal scheduling of batteries that perform reactive balancing [Smets et al., 2023]. Nevertheless, this line of work is tangent to our analysis, which is focused on the *design* of balancing markets.

A similar analysis on the effect of uncoordinated regulation in spatially integrated electricity markets has been conducted by Bushnell in the context of carbon reduction policies [Bushnell et al., 2017]. The authors conclude that regulatory interventions can lead to a distortion of the merit order, which may in turn lead to inefficiencies.

This chapter completes and extends the model proposed in [Papavasiliou

and Bertrand, 2021] in order to assess the back-propagation induced by different imbalance and balancing pricing schemes. The modelling and theoretical contributions can be stated as follows: (i) we characterize an equilibrium for the "adder on BRPs" design, which allows us to abandon the agent-based modeling method used in [Papavasiliou and Bertrand, 2021], (ii) we provide a novel analysis of a newly proposed pricing scheme, the "adder on BRPs and BSPs" design, and (iii) we apply our analysis to a cross-border setting. From a policy standpoint, our analysis shows the inability of both the "adder on BRPs" and the "adder on BRPs and BSPs" designs to support an optimal dispatch in a cross-border setting, in contrast to the "RT market for reserve" design which can achieve this objective.

The remainder of the chapter is structured as follows. Section 3.2 goes on to describe the functioning of European balancing markets and balancing platforms, and introduces our notation and model. Sections 3.3 and 3.4 describe the optimal strategies of agents under the different designs analyzed in this work, as well as the resulting market equilibrium. Section 3.5 illustrates and compares the market equilibria in a two-zone setting. Section 3.6 concludes.

3.2 European Balancing Market

This section presents a single-zone balancing market and then introduces balancing platforms for representing cross-zonal integration in balancing operations.

3.2.1 Single-Zone Balancing Market

The functioning of European balancing markets is outlined in the *Electricity Balancing Guideline* (EBGL) and described in a stylised manner hereunder.

TSOs are responsible for the operational security of the grid. They hold reservation auctions for ensuring an adequate level of available reserve capacity of different types in real time. Balancing capacities can be differentiated according to their activation time and include *frequency containment reserve* (FCR), *automatic frequency restoration reserve* (aFRR), *manual frequency restoration reserve* (mFRR), and *replacement reserve* (RR). The discussion in the paper is targeted at manual frequency restoration reserve that is dispatched through balancing energy auctions. Replacement reserve and automatic frequency restoration reserve are also dispatched through balancing energy auctions, but they are ignored in order to highlight the effect of introducing adders on the interaction between two cash flows with comparable settlement timeframes: mFRR balancing energy and imbalance. Any subsequent reference to balancing capacity and energy in this chapter will respectively refer to the mFRR capacity available for activation by the TSOs and to the mFRR capacity dispatched by the TSOs.

BRPs (flexibility consumers) are owners of portfolios that consist of residential, commercial and industrial load, as well as generation assets. According to the EBGL, they shall strive to be balanced or to help the system be balanced (article 17.1 of EBGL) and they are financialy responsible for their imbalance (article 17.2 of EBGL). Their imbalance relative to their ex-ante position is charged at the *imbalance price*. For the sake of this analysis, BRPs can be considered as price-inelastic energy bids.

BSPs are flexibility providers that participate in the balancing energy auctions. They belong to a BRP portfolio, and they include a wide range of assets, such as classical thermal units (CCGT, OCGT, ...), battery aggregations, and industrial and/or commercial demand response. BSPs can offer various reserves, depending on their characteristics and on the qualification criteria set by the TSO. BSPs can be considered as elastic suppliers of balancing energy in the context of our models.

The term "balancing the market" refers to the process whereby a TSO activates balancing energy from BSPs, in order to cover the aggregation of the BRPs' inelastic imbalance. We proceed now with a description of the balancing process. We have voluntarily left aside the reserve procurement auctions, as their representation is not required in order to highlight the issues that emerge from pricing asymmetries resulting from adders.

Firstly, BSPs submit their balancing energy bids to a balancing energy auction which is organised by the TSO. The balancing energy auction is assumed to clear at a uniform price, following the pricing scheme of the European balancing platforms [ACER, 2020a].

Secondly, the aggregation of the BRPs' inelastic imbalance is revealed. The TSO clears the balancing energy auction in order to balance the market and a *platform price*, λ_P, is generated as the marginal cost of balancing energy activation. Given the available information, one can also compute a *scarcity component price*, λ_R. This represents the value of balancing capacity at the time of clearing and can be set by an ORDC.

Between the first and second step, BSPs, as part of a BRP portfolio, can decide to perform reactive balancing and self-activate their assets. In this case, the activated energy is considered part of the BRP's imbalance and is charged at the imbalance price.

The ISHM and the pricing methodology leave the door open for the introduction of scarcity components in the balancing price, λ_{bal}, and the imbalance price, λ_{imb}. These prices are used respectively for remunerating BSPs' balancing energy and for settling BRPs' imbalance.[1] The application of such adders can result in different designs, as shown in Table 3.1. The default design is the "no adder" policy, where the balancing and imbalance price are equal to the platform price. The "adder on BRPs" and "adder on BRPs and BSPs" designs introduce an adder on the imbalance price and on the imbalance and balancing price, respectively. Finally, the "RT market for reserve" design has an adder

[1] This analysis does not include boundary cases where the price cap for the balancing energy market and the imbalance settlement is reached.

Table 3.1: Balancing and imbalance prices under the various designs that are debated in European balancing market design.

	λ_{bal}	λ_{imb}	Res. price
No adder	λ_P	λ_P	0
Adder on BRPs	λ_P	$\lambda_P + \lambda_R$	0
Adder on BSPs and BRPs	$\lambda_P + \lambda_R$	$\lambda_P + \lambda_R$	0
RT Market for reserve	$\lambda_P + \lambda_R$	$\lambda_P + \lambda_R$	λ_R

on the imbalance and balancing price, and additionally trades balancing capacity in real time. In this last design, the balancing capacity that has not been activated is entitled to the real-time reserve price, which is equal to an adder computed from an ORDC. More specifically, the "RT market for reserve" design proposes to introduce a market for *balancing capacity imbalance* which is equivalent to a market for balancing capacity that is conducted in real time. The reader is referred to section 1.3.1 for additional details on the functioning of RT market for reserve.

3.2.2 Cross-Border Balancing Platforms

The transition from one zone to multiple zones requires cross-border coordination, which is the goal of the European balancing platforms. These platforms aim at coordinating the dispatch of balancing energy from different zones, and are called PICASSO (for aFRR) and MARI (for mFRR). Their objective is to cover the TSOs' demand, at least cost, by activating balancing energy from the BSPs of multiple zones. They have gone live in 2022 and are operating in Germany, Austria and the Czech Republic. The other European TSOs are expected to join the platforms in 2024 or 2025. MARI clears every 15 minutes and PICASSO clears every 4 seconds.[2]

TSOs that are connected to the platforms first receive the balancing bids from the BSPs. They filter these bids in order to suppress the ones that could create congestion and transmit the others to the platform. Afterwards, they send their demand for the activation of balancing energy to the platform[3]. The platform clears a balancing energy auction and informs TSOs on which bids of their control area have been accepted. Finally, TSOs inform BSPs the bids of which are accepted. TSOs are also responsible for the settlement between the

[2]MARI can clear more than one time per 15 minutes at the TSOs' request.

[3]In practice, MARI accepts elastic demand bids from TSOs, but we restrict our investigation to inelastic demand.

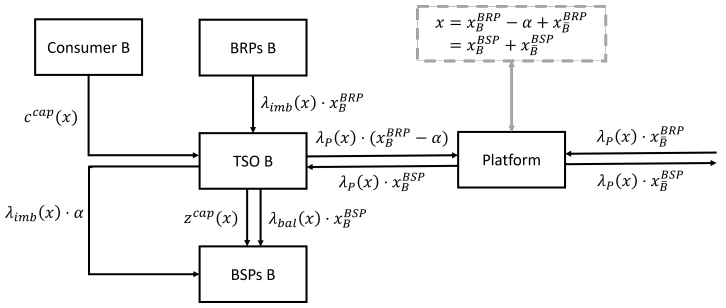


Figure 3.1: Cash flow over multiple zones for a general European cross-border balancing market.

BRPs and BSPs and the platform. This balancing process, and the cash flow between the different agents, is represented graphically in Fig. 3.1 for the case of two zones: zone B and zone $\bar{B}$, where zone $\bar{B}$ corresponds to the rest of the system.

The demand for activation of balancing energy in zone i is denoted as x_i^{BRP} if there is no reactive balancing and as $x_i^{BRP} + \alpha$ in the case with reactive balancing. Note that the TSO cannot distinguish between the inelastic imbalance from BRPs and reactive balancing. The demand for balancing energy accounts for all the imbalances generated by agents mobilizing flexiblity outside of the balancing energy auction. This includes demand response aggregators not participating in the balancing energy auctions. The activated balancing energy in zone i is denoted as x_i^{BSP}. Settlements between the platform and the TSOs are charged at the platform price cleared by the balancing energy auction, but TSOs are free to unilaterally introduce adders, corresponding to the designs proposed in table 3.1, for the settlements with the BRPs and BSPs.

These potential pricing asymmetries between the platform price and the balancing and imbalance price might result in a capacity cost component c^{cap} borne by all the consumers in the zone introducing the adder, which is socialized through the grid tariff. Outside of balancing settlement, BSPs are also entitled to a capacity settlement z^{cap} for their unused balancing capacity if their zone operates a "RT market for reserve". Both of these components are described for the different designs in Table 3.2.

Note that all $c^{cap}(x)$, except for the "no adder" one, should include a self-activation component α but they were dropped due to the level of self-activation at equilibrium being equal to 0 for the "adder on BSPs and BRPs" and the "RT market for reserve". These results are shown in section 3.4.

Table 3.2: Cost of capacity and capacity settlement

	$c^{cap}(x)$	$z^{cap}(x)$
No adder	0	0
Adder on BRPs	$-\lambda_R(x) \cdot (x_B^{BRP} - \alpha)$	0
Adder on BSPs and BRPs	$\lambda_R(x) \cdot (x_B^{BSP} - x_B^{BRP})$	0
RT Market for reserve	$\lambda_R(x) \cdot (P^{max} - x_B^{BRP})$	$\lambda_R(x) \cdot (P^{max} - x_B^{BSP})$

3.3 Optimal Decisions of Agents

The objective of the analytical model is to identify the optimal strategy of risk-neutral BSPs that maximize profit.[4] BSPs are assumed to participate in a single-product balancing market with a uniform balancing price and an imbalance price which is based on the balancing price. The agents can either participate in the energy balancing auction or perform reactive balancing. BRPs are assumed to be inelastic. We will first describe the profit function of a fringe BSP with P^+ MW of upward balancing capacity and a marginal cost of C €/MWh. We will then derive its optimal strategy and discuss the impact of BSPs' level of information on the demand for balancing energy on their bidding strategy. The notation and analysis presented here are an extension of the model presented in [Papavasiliou and Bertrand, 2021].

The model has been kept simple on purpose to highlight the effects brought forth by the introduction of adders. Inefficiencies encountered in a simple stylized model are symptoms of underlying design flaws. In this spirit, well-documented sources of distortion or inefficiencies in electricity markets such as market power [Borenstein et al., 2002, Prabhakar Karthikeyan et al., 2013], collusive bidding by suppliers [Ocker et al., 2018], arbitrage opportunities between different products [Oren, 2001], gaming opportunities created by congestion management actions [Chaves-Ávila et al., 2014], and frictions resulting from ramping or commitment constraints [Petitet et al., 2019] are ignored. Nevertheless, the modelling framework is quite flexible, and additional components can be introduced in the analysis. For example, the introduction of reservation cost and a day-ahead balancing capacity market is discussed in the appendix of the chapter.

The focus on upward balancing capacity is justified by the fact that scarcity adders are expected to be activated when the system is under stress due to insufficient generation, i.e. when it is short. BSPs with downward balancing capacity will not be affected by the introduction of scarcity adders as such

[4] Market agents may exhibit risk-averse behaviour in practice.

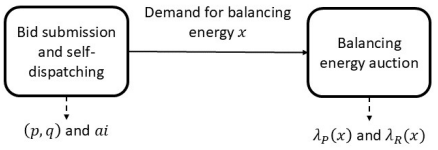


Figure 3.2: Two-stage balancing process

adders will be zero when these BSPs are dispatched.[5]

3.3.1 Balancing and Imbalance Payoffs of Agents

The two-stage balancing process is represented in figure 3.2. In the first stage, the agent submits a price-quantity bid (p, q) to the energy balancing auction and decides on its level of reactive balancing ai. Note that the sum of the level of reactive balancing and of the bid quantity must be lower than the total capacity of the agent. In the second stage, the TSO dispatches balancing energy through the balancing energy auction in order to cover the system imbalance x, which is the demand for balancing energy. This demand is assumed to be drawn from a known distribution X with cumulative distribution function $F(\cdot)$ and probability density function $f(\cdot)$. The resulting price from the balancing energy auction, $\lambda_P(x)$, is a function of the random demand for balancing energy. It can be defined as the price offer of the most expensive accepted energy bid. Alternatively, the platform price is equal to the dual variable of the market clearing constraint of the economic dispatch problem that is solved by the system operator in order to balance the system.

The scarcity component, $\lambda_R(x)$, is obtained through an operating reserve demand curve. Scarcity pricing based on ORDC adders takes the leftover capacity in the system as an input, but leftover capacity can be equivalently expressed as a function of the demand for balancing energy by assuming that the leftover capacity in the system is the total capacity in the system minus the inelastic energy demand. The scarcity component is a non-decreasing function of demand.

The price-quantity bid and the decision to self-dispatch depends on the expectation of agents regarding their payoff which is a function of the platform price, the scarcity component and the distribution of the demand for balancing energy. It is important to note that decisions are taken before the revelation of the demand for balancing energy.The balancing payoff for a BSP, given a random demand for balancing energy, is the uniform balancing price multiplied by the quantity bid if the price bid is lower than the platform price, or 0 if the bid is not accepted.

[5] The paper focuses on upward reserve but the rise in solar generation and the resulting steep increasing ramp in the morning, the so-called "duck curve", has increased the need for downward reserve. Further research is required to asses how the mechanism can accommodate a scarcity adder on downward reserve.

$$z_B(p,q,x) = \begin{cases} \lambda_{bal}(x) \cdot q & \text{if } p \leq \lambda_P(x), \\ 0 & \text{else} \end{cases} \tag{3.1}$$

Note that bids are assumed to be either fully selected, if they are at-the-money or in-the-money, or not selected at all. The expectation of this payoff can be reformulated as follows, with the operator $\mathbb{E}_X$ being the expectation over the random variable X:

$$z_B(p,q) = \mathbb{E}_X[z_B(p,q,\cdot)] = \int_{\lambda_P(x) \geq p} (\lambda_{bal}(x) - C) dF(x) \cdot q \tag{3.2}$$

The payoff of reactive balancing is found by first deriving the level of reactive balancing performed by an agent. This is found by solving the following optimization problem:

$$\max_{ai} \quad (\mathbb{E}_X[\lambda_{imb}] - C) \cdot ai \tag{3.3}$$

$$s.t. \quad ai + q \leq P^+ \tag{3.4}$$

$$ai \geq 0 \tag{3.5}$$

If $C \geq \mathbb{E}_X[\lambda_{imb}]$, the optimal level of reactive balancing $ai^\star$ is 0, else $ai^\star$ is equal to the leftover capacity from the balancing energy auction. The reactive balancing payoff is then described as follows:

$$z_I(q) = \begin{cases} (\mathbb{E}_X[\lambda_{imb}] - C) \cdot (P^+ - q) & \text{if } C \leq \mathbb{E}_X[\lambda_{imb}], \\ 0 & \text{else.} \end{cases} \tag{3.6}$$

The optimal strategies for the "no adder", "adder on BRPs" and "adder on BRPs and BSPs" designs are then found by maximizing the sum of the balancing and of the reactive balancing payoff.

A real-time market for balancing capacity requires the introduction of a new component in the payoffs. The unused balancing capacity of BSPs is now remunerated by the reserve price. As the balancing and imbalance prices are equal to the platform price plus the reserve component in the "RT market for reserve" design (see table 3.1), the real-time payoff can be reformulated as follows for a random demand x, with $z_I(ai,x)$ being the reactive balancing payoff for self-activating ai MWh:

$$\begin{aligned} & z_B(p,q,x) + z_I(ai,x) \\ &= \begin{cases} (\lambda_P(x) + \lambda_R(x) - C) \cdot (q + ai) + \lambda_R(x) \cdot (P^+ - q - ai) \\ \qquad \text{if } \lambda_P(x) \leq p, \\ (\lambda_P(x) + \lambda_R(x) - C) \cdot ai + \lambda_R(x) \cdot (P^+ - ai) \\ \qquad \text{else,} \end{cases} \qquad (3.7) \\ &= \begin{cases} (\lambda_P(x) - C) \cdot (q + ai) + \lambda_R(x) \cdot P^+ & \text{if } \lambda_P(x) \leq p, \\ (\lambda_P(x) - C) \cdot ai + \lambda_R(x) \cdot P^+ & \text{else.} \end{cases} \qquad (3.8) \end{aligned}$$

This allows us to rewrite both the balancing and imbalance payoff as a function of the platform price:

$$z_I(q) = \begin{cases} (\mathbb{E}_X[\lambda_P] - C) \cdot (P^+ - q) & \text{if } C \leq \mathbb{E}_X[\lambda_P] \\ 0 & \text{else,} \end{cases} \tag{3.9}$$

$$z_B(p, q) = \int_{\lambda_P(x) \geq p} \big((\lambda_P(x) - C) \cdot q\big) dF(x) + \mathbb{E}_X[\lambda_R] \cdot P^+. \tag{3.10}$$

The objective of this reformulation is to isolate the scarcity component remuneration $\mathbb{E}_X[\lambda_R] \cdot P^+$ from the standard imbalance and balancing payoff, in order to highlight the correspondence between the payoffs of the "RT market for reserve" and "no adder" designs up to a constant.

3.3.2 Optimal Balancing Market Bid

The strategy for characterizing the optimal behaviour under the different designs is an extension of [Papavasiliou and Bertrand, 2021] where we additionally analyze the "adder on BRPs and BSPs design". Only the statement and a brief intuition of the results are provided here. The complete proofs are available in the appendix. These proofs consider fringe agents reacting to an exogenous platform price. The platform price is also assumed to be strictly monotonic increasing. The inverse of the platform price, i.e. the supply function, is assumed to be differentiable almost everywhere.[6] Section 3.4 shows that these assumptions hold at equilibrium.

Proposition 3.1 (Bidding Strategy – No Adder)**.** *The optimal strategy for a fringe agent under a "no adder" design is to bid truthfully in the balancing auction.*

Bidding more or less than the marginal cost in the balancing energy auction will result in a lower balancing payoff for an agent, as the agent can respectively lose some potential payoff (in the case of overbidding) or be unprofitable (in the case of underbidding). One can then conclude that it is optimal for every agent to bid its full capacity in the balancing auction, since the payoff of the balancing auction will be (i) equal to the payoff from self-activation whenever the agent is in the money in the balancing auction, and (ii) is higher than the payoff of self-activating whenever the agent is out of the money.

Proposition 3.2 (Bidding Strategy – Adder on BRPs)**.** *The optimal strategy for a fringe agent under an "adder on BRPs" design is to bid truthfully in the balancing market if*

$$C \geq \mathbb{E}_X[\lambda_P + \lambda_R] - \int_{\lambda_P(x) \geq C} (\lambda_P(x) - C) dF(x),$$

else to perform reactive balancing with its full capacity.

[6] At equilibrium, the supply function may exhibit breaking points due to BSPs resorting to reactive balancing. This can cause the supply function to not be differentiable everywhere.

The pricing asymmetry of this design can incentivize BSPs to self-activate when they have a low marginal cost. If a BSP is very likely to be activated, the little it would lose when the imbalance price is not sufficient to cover its cost can be compensated by the additional payoff of the imbalance settlement, compared to the balancing energy auction, due to the scarcity component in the former.

In order to analyze the next design, we define $(\lambda_P + \lambda_R)^{-1}$ as the inverse function of the sum of the platform price and the scarcity component under the adder on BRPs and BSPs design. This sum is equal to both the balancing and imbalance price (see Table 3.1). The expression $(\lambda_P + \lambda_R)^{-1}(p)$ is then the level of demand for balancing energy such that the balancing price p is attained. Note that $\lambda_P + \lambda_R$ has a well-defined inverse everywhere because the platform price is strictly monotonic increasing and the reserve component is non-decreasing.

Proposition 3.3 (Bidding Strategy – Adder on BRPs and BSPs)**.** *The optimal strategy for a fringe agent under an "adder on BRPs and BSPs" design is to bid its full capacity in the balancing energy market at price*

$$\lambda_P((\lambda_P + \lambda_R)^{-1}(C)).$$

BSPs should internalize the value of the adder in their balancing energy bid so as to ensure that they are always activated when the balancing price is higher than or equal to their marginal cost. This corresponds to bidding the platform price $\lambda_P(x')$ for x' such that

$$(\lambda_P + \lambda_R)(x') = C.$$

Proposition 3.4 (Bidding Strategy – RT Market for Reserve)**.** *The optimal strategy for a fringe agent under a RT market for reserve design is to bid truthfully in the balancing energy market.*

The profit function of the "RT market for reserve" design is equal to the one of the "no adder" design up to a constant $\mathbb{E}_X[\lambda_P] \cdot P^+$, independent of the agent's strategy.

The optimal bidding strategies for a BSP under the different designs are not modified by the transition to multiple zones. As long as (i) the platform price and the reserve component price as a function of the zonal demands for balancing energy and (ii) the probability measures of these demands are known, the optimal strategies described above remain valid.

3.3.3 Level of Information Regarding the Demand for Balancing Energy

In practice, one could argue that BSPs have more freedom concerning the self-activation of their plant than what has been described earlier. Current Belgian bidding rules require BSPs to bid in the balancing energy auction and allow

them to retract their bid up to 25 minutes before the clearing of the balancing auction. Under this regulation, BSPs can decide to self-activate their plant only when the state of the system at time t minus 25 guarantees them a favorable outcome by self-activating. If the state of the system is unfavorable, they can opt to bid in the balancing market instead, and thus avoid being exposed to any risk.

From a modelling standpoint, we can introduce information on the demand for balancing energy by using a random variable y drawn from a known distribution Y with PDF $g(\cdot)$ and CDF $G(\cdot)$. The random variable y is assumed to be observable at the moment when the BSP is called to reach a decision about reactive balancing, e.g. it can include the realized imbalance of the previous interval, wind forecasts, load forecasts, etc. This random variable thus provides information that is revealed to the BSPs when submitting their energy bid. BSPs can then consider whether to self-activate their assets or participate in the balancing auction, depending on the distribution of the demand given y, $X|y$, its CDF $F(\cdot|y)$, and the platform price, balancing price, and imbalance price for a given demand given y, $\lambda_P(x|y)$, $\lambda_{bal}(x|y)$ and $\lambda_{imb}(x|y)$.

The balancing payoff of an agent can then be expressed as follows:

$$z_B(p,q,y) = \int_{\lambda_{bal}(x|y)\geq p} (\lambda_{bal}(x|y)\cdot q) dF(x|y) \tag{3.11}$$

The payoff from reactive balancing as a function of the information is found by solving the following optimization problem:

$$\max_{ai(y)} \quad \int\int (\lambda_{imb}(x|y) - C)\cdot ai(y)) dF(x|y) dG(y) \tag{3.12}$$

$$s.t. \quad ai(y) + q(y) \leq P^+ \tag{3.13}$$

$$ai \geq 0 \tag{3.14}$$

Notice that both the balancing and imbalance profit functions are separable for y, meaning that they are parametrized by y but do not couple different values of y with each other. This allows us to model the level of information by combining the optimal strategy of agents for different distributions of demand for balancing energy without information and to fall back to the basic setting.

3.4 Market Equilibrium

This section commences by characterizing the Nash equilibria resulting from the optimal strategy outlined earlier in a single zone. These equilibria correspond to the ones that would emerge in a balancing market that is not connected to a cross-border platform. We then extend the analysis to multiple zones and we discuss the ensuing inefficiencies.

The result presented here assumes a truthful merit order curve $MC(x)$, which is strictly monotonic increasing and differentiable, as well as a system

capacity P^{max} which is greater than the upper bound of the support of the distribution of the random demand for balancing energy.

3.4.1 Single-zone

Proposition 3.5 (Equilibrium – "No Adder" and "RT market for reserve")**.** *The Nash equilibrium generated by fringe agents under the "no adder" and "RT market for reserve" designs is characterized by all agents participating truthfully in the balancing energy auction and the following platform price:*

$$\lambda_P(x) = MC(x).$$

Proof. The agents' optimal strategies consisting of bidding truthfully are independent from the other agents' bidding behavior. This behavior, coupled with the balancing energy auction selecting bids in increasing price order, results in the platform price following the merit order, and being strictly monotone increasing and differentiable. This confirms the validity of propositions 3.1 and 3.4. □

We now define $\lambda_P(x, \alpha)$ as the platform price for energy demand x, a total of α MWh of reactive balancing from the cheapest BSPs with upward balancing capacity, and we consider what happens when other BSPs bidding truthfully. The platform price in this situation is as follows:

$$\lambda_P(x, \alpha) = \begin{cases} MC(x - \alpha) \text{ if } x < \alpha, \\ MC(x) \text{ if } x > \alpha, \\ \text{price indeterminacy between } MC(0) \text{ and } MC(\alpha) \text{ else.} \end{cases} \tag{3.15}$$

If $x < \alpha$, $x - \alpha$ MWh of downward balancing capacity has to be activated in order to balance the excessive self-activation by the agents, resulting in $\lambda_P(x, \alpha) = MC(x - \alpha)$. If $x > \alpha$, there is no price distortion and $\lambda_P(x, \alpha) = MC(x)$. If $x = \alpha$, no balancing energy is activated through the balancing energy auction and there is a price indeterminacy.

In a multi-zone setting, the reserve component in a zone is impacted by the level of reactive balancing in that zone contrarily to the single-zone setting where the presence of downward balancing capacity allows us to ignore it. To see this, note that if the level of reactive balancing is greater than the demand for balancing energy, then the potential curtailment of downward balancing capacity that was dispatched to cover the excessive reactive balancing can be assimilated as upward balancing energy. In case of an increased demand for balancing energy, reducing the level of the dispatched downward balancing capacity will contribute towards reducing the imbalance and will therefore not impact the level of upward balancing capacity in a single zone setting where only that zone can provide upward balancing capacity. If several zone can be

dispatched to cover reactive balancing, the potential curtailment of downward balancing capacity in another zone cannot be assimilated as upward balancing energy.

The opportunity cost of participating in the balancing auction given a level α of self-balancing from the cheapest agents with upward balancing capacity for an agent with marginal cost C is:

$$\begin{aligned} z(\alpha, C) =& (\mathbb{E}_X[\lambda_P(\cdot, \alpha) + \lambda_R(\cdot)] - C) \\ & - \int_{\lambda_P(x,\alpha) \geq C} (\lambda_P(x, \alpha) - C) dF(x). \end{aligned} \tag{3.16}$$

If $z(\alpha, C) < 0$, an agent with marginal cost C should bid truthfully in the balancing auction for a level α of reactive balancing. If $z(\alpha, C) > 0$, the agent should self-activate its capacity. Note that the reserve component is independent from the level of reactive in this single zone setting. To see this, note that if the level of reactive balancing is greater than the demand for balancing energy, then the potential curtailment of downward balancing energy that was dispatched to cover the excessive reactive balancing can be assimilated as upward balancing capacity. Reducing the level of the dispatched downward balancing capacity will contribute towards reducing the imbalance and will therefore not impact the availability of upward balancing capacity.

Proposition 3.6 (Equilibrium – Adder on BRPs)**.** *If $z(\alpha, MC(\alpha))$ is continuous, there exists a unique Nash equilibrium generated by fringe agents under the "adder on BRPs" design characterized by an equilibrium level of reactive balancing, α^*, such that $0 \leq \alpha^* \leq P^{max}$, and with other BSPs bidding truthfully. This optimal level of reactive balancing is equal to (i) 0 if $z(0, MC(0)) < 0$, (ii) P^{max} if $z(P^{max}, MC(P^{max})) > 0$ or (iii) α^* characterized by the identity*

$$z(\alpha^*, MC(\alpha^*)) = 0. \tag{3.17}$$

This equilibrium level of reactive balancing generates platform prices equal to $\lambda_P(x, \alpha^*)$.

The existence of an equilibrium with $0 < \alpha^* < P^{max}$ relies on on the continuity of z. The stability of α^* is derived analytically. Stability in this context refers to a level of reactive balancing for which no agent has an incentive to deviate from its decision. BSPs after α^* on the merit order prefer to participate in the balancing auction and agents before α^* prefer to resort to reactive balancing. The uniqueness of the equilibrium results from the monotonicity of z. The complete proof can be found in the appendix.

Note that assuming a positive distribution for the demand for balancing energy results in z being continuous. Under this assumption, the probability of a particular demand occurring is infinitesimal. Discrete random demand can generate a price indeterminacy if the level of reactive balancing is equal to the imbalance. This breaks the continuity of z and an example of a system without a pure-strategy equilibrium can be found in the appendix.

Proposition 3.7 (Equilibrium – Adder on BRPs and BSPs)**.** *If $MC(x)-\lambda_R(x)$ is strictly monotonic increasing, there exists a Nash equilibrium generated by fringe agents under an "adder on BRPs and BSPs" design. It is characterized by all agents participating in the balancing energy auction and internalizing the value of the adder in their balancing energy bid. The produced platform price is described as follows:*

$$\lambda_P(x) = MC(x) - \lambda_R(x).$$

Proof. The agents' optimal strategy is to bid at their marginal cost minus the scarcity component. This bidding behavior, combined with the balancing energy auction selecting bids in increasing price order, results in the platform price following the merit order minus the scarcity component and being strictly monotonic increasing and differentiable. This confirms the validity of proposition 3.3. □

If $MC(x)-\lambda_R(x)$ is not strictly monotonic increasing, the optimal strategy derived in proposition 3.3 could modify the order of activation.

In terms of efficiency, the "no adder", "adder on BRPs and BSPs" and "RT market for reserve" designs support the optimal dispatch for a single zone, as they do not modify the order of activation specified by the truthful merit order. The "adder on BRPs" design increases the cost by inducing the dispatch of assets out of the merit order.

3.4.2 Multiple Zones

The characterization of an equilibrium in a setting with multiple zones requires introducing an aggregation operator $\cup$ for the aggregation of offer curves from different zones. Given $B_i(q)$, the offer curve in zone i, the aggregated offer curve, $B(q)$, can be obtained through the aggregation operator, as follows:

$$B(q) = \cup_i B_i(q) = \{\pi : B_i(q_i) = \pi \text{ for all } i \text{ and } \sum_i q_i = q\}. \tag{3.18}$$

The optimal strategies derived in section 3.3 remain valid in a multi-zone setting, and are used in order to derive offer curves under different designs, as shown in Table 3.3. $\lambda_{R,i}$ is the reserve demand curve in zone i and α_i is the optimal level of self-activation in zone i.

For the "adder on BRPs" design, the opportunity cost function has to be modified in order to account for multiple zones. The assumption regarding the scarcity component not being impacted by the level of self-activation needs to be revisited. Excessive self-activation in a multi-zone setting is covered by activating downward balancing capacity from all zones. This reduces the total level of available upward balancing capacity in the zone with self-dispatched assets. This means that we need to define the scarcity component as a function of both the level of aggregated demand for balancing energy over all zones,

Table 3.3: Offer curves under different designs.

Design in zone i	$B_i(x)$
No adder	$MC_i(x)$
Adder on BRPs and BSPs	$MC_i(x) - \lambda_{R,i}(x)$
Adder on BRPs	$\begin{cases} MC_i(x-\alpha_i) & \text{if } x \leq \alpha_i \\ MC_i(x) & \text{else} \end{cases}$
RT market for reserve	$MC_i(x)$

as well as the level of self-activation in the zone with the "adder on BRPs", $\lambda_R(x, \alpha)$, and to update the opportunity cost of self-activation, as follows:

$$z(\alpha, C) = (\mathbb{E}_X[\lambda_P(\cdot, \alpha) + \lambda_R(\cdot, \alpha)] - C) - \int_{\lambda_P(x,\alpha) \geq C} (\lambda_P(x, \alpha) - C) dF(x), \tag{3.19}$$

This modifies the condition for an equilibrium level of self-activation and might lead to multiple equilibria if $z(\alpha, MC(\alpha))$ is not strictly monotonic decreasing in α.

Two conclusions can be drawn from the aggregation of the offer curves presented in table 3.3. First, only the introduction of a "RT market for reserve" does not affect the optimal dispatch. Both the "adder on RBPs and BSPs" and the "adder on BSPs" modify the bidding incentives in the zone implementing an adder, and result in a suboptimal aggregated offer curve. Second, the suboptimal aggregated offer curves generate lower platform prices than the one generated by the aggregation of the truthful merit order curves.

3.5 Illustration on a Stylized Example

The examples presented in this section assume a maximum level of upward balancing capacity P^{max}, and a BSP merit order curve $MC(x)$ which is a function of the level of demand for balancing energy, x. The demand is drawn from a known distribution X with CDF $F(\cdot)$. The scarcity component λ_R is obtained from an operating reserve demand curve defined as a function of the level of demand for balancing energy in the system.

This section presents four examples: (i) a single-zone example without information on the level of demand for balancing energy in the system, (ii) a single-zone example with information on the level of demand for balancing energy, (iii) a two-zone example without cross-border congestion and with information on the level of demand for balancing energy, and (iv) a two-zone example with cross-border congestion and with information on the level of demand for balancing energy.

Table 3.4: Example 3.1 – Expected platform and scarcity prices, level of reactive balancing and activation cost for the different designs

	No adder	Adder on BRPs and BSPs	Adder on BRPs	RT market for reserve
$\mathbb{E}_X[\lambda_P]$ (€/MWh)	60.00	55.83	60.00	60.00
$\mathbb{E}_X[\lambda_R]$ (€/MWh)	0.00	4.17	4.17	4.17
α (MWh)	0.00	0.00	0.00	0.00
Activation cost (€)	833.38	833.38	833.38	833.38

3.5.1 Example 3.1: Single Zone without Information on the Demand for Balancing Energy

In this example, we assume that the demand is uniformly distributed between -100 MWh and 100 MWh and that the merit order curve is described as follows:

$$MC(x) = x/2 + 60 \quad €/\text{MWh}. \tag{3.20}$$

The scarcity price component is defined as

$$\lambda_R(x) = \begin{cases} 0 \ €/\text{MWh} & \text{if } x \leq 0, \\ x/6 \ €/\text{MWh} & \text{else}, \end{cases} \tag{3.21}$$

which can be equivalently formulated as a function of the leftover capacity in the system, $\lambda_R^r(r)$, assuming a maximum level of balancing capacity in the system P^{max}, equal to 200 MW in our case.

$$\lambda_R^r(r) = \lambda_R(P^{max} - r) = \begin{cases} (P^{max} - r)/6 & \text{if } r \leq P^{max}, \\ 0 & \text{else}. \end{cases} \tag{3.22}$$

All BSPs (i) bid truthfully under the no-adder and RT market for reserve design, thus $\lambda_P(x) = MC(x)$ (see proposition 3.5); (ii) bid at their marginal cost minus the level of the adder at their position on the merit order under the adder on BRPs and BSPs design, thus $\lambda_P(x) = MC(x) - \lambda_R(x)$ (prop. 3.7); (iii) bid in the balancing energy auction at their marginal cost under the adder on BRPs design, thus $\lambda_P(x) = MC(x)$ (prop. 3.12). No BSP does reactive balancing, as the opportunity cost of the cheapest generator when no asset is self-activating is negative. If participating in the balancing auction is more profitable for the cheapest generator, then this is also the case for every generator.

Table 3.4 presents the expected platform price, the expected scarcity component, the level of self-activation, and the cost of reserve activation under the four designs. The four designs result in the same activation cost as the merit order is not distorted.

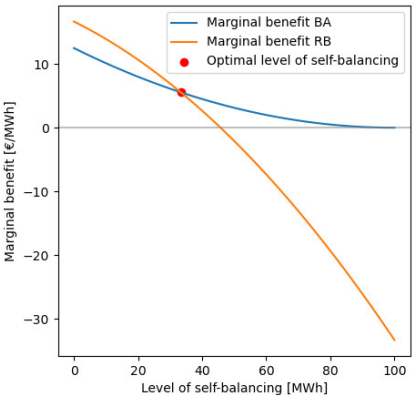


Figure 3.3: Comparison of marginal benefit of reactive balancing and balancing auction for the frontier agent in the case of example 3.2.

3.5.2 Example 3.2: Single Zone with Information on the Demand for Balancing Energy

In this example, BSPs have some information on the level of demand that the system will be exposed to, and are specifically aware of the sign of the required balancing activation. Specifically, we assume a draw with a probability 0.5 of having a negative demand that is distributed uniformly between -100 MWh and 0 MWh and a probability 0.5 of having a positive demand that is distributed uniformly between 0 MWh and 100 MWh.

All the parameters are identical to the previous example. The BSPs' strategy under the "no-adder", "adders on BRPs and BSPs" and "RT market for reserve" designs are not modified by the introduction of information on the imbalance, but there is an impact on the "adder on BRPs" design. The optimal level of reactive balancing in the "adder on BRPs" design is (i) 0 MWh if the system imbalance is in the interval $[-100, 0]$ MWh, and (ii) 33.33 MWh if the system imbalance is in the interval $[0, 100]$ MWh. The optimal level of reactive balancing is found by resolving the identity of Eq. (3.17) for α, the level of self-activation. In the first interval, no generator self-balances, as $z(\alpha, MC(\alpha)) \leq 0$ for all α. For the second interval, $\alpha = 33.33$ MWh does satisfy the identity. This process is illustrated graphically in Fig. 3.3 by splitting the opportunity cost between the balancing auction and reactive balancing component. This figure represents the difference in profits for the frontier agent (i.e. the last agent to self-activate) for different levels of reactive balancing.

The platform prices for the four designs are presented in Fig. 3.4 as a function of the level of demand for balancing energy. Agents bid truthfully under the "no adder" and "RT market for reserve" designs, and they internalize

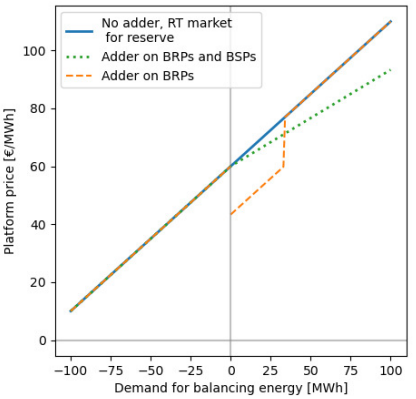


Figure 3.4: Offer curve in example 3.2.

Table 3.5: Example 3.2 – Impact of additional information on the demand for balancing energy in the market equilibrium

Add. information	With			Without
Imbalance interval	[-100; 0]	[0; 100]	[-100; 100]	[-100; 100]
$\mathbb{E}_X[\lambda_P]$ (€/MWh)	35.00	79.42	57.21	60.00
$\mathbb{E}_X[\lambda_R]$ (€/MWh)	0.00	8.33	4.17	4.17
α (MWh)	0.00	33.33	16.67	0.00
Activation cost (€)	-2166.63	3925.99	879.68	833.38

the reserve adder under the "adder on BRPs and BSPs" design. Some of the BSPs decide to resort to reactive balancing if they know that the demand will be between 0 MWh and 100 MWh under the "adder on BRPs". This self-activation results in a translation of the merit order curve for negative balancing activation up to the level of reactive balancing.

The metrics concerning both intervals are presented in the first two columns of Table 3.5. Columns 3 and 4 compare the result for the "adder on BRPs" design with and without additional information on the demand for balancing energy. The reactive balancing results in an inefficient dispatch that increases the total activation cost of the system. It also decreases the platform price, which is beneficial to the BRPs.

3.5.3 Example 3.3 – Two Zones without Cross-Border Congestion and with Information on the Demand for Balancing Energy

We now refer to the system mentioned in examples 3.1 and 3.2 as zone B, and connect it to a new zone with an unlimited interconnector capacity between the two zones. The system in this new zone, called zone D, is four times larger than the one in zone B, resulting in a less steep merit order curve (see Eq. (3.23)).

$$MC_D(x) = x/8 + 60 \quad €/\text{MWh} \tag{3.23}$$

The example is intended to mimic, in a highly stylized setting, the interaction between Belgium and Germany, hence the initials of the zones. We limit the exposition to the 2-zone case with zone D implementing a "no adder" policy in order to analyze the effect of the pricing asymmetry between the zones, since this has also dominated the policy discussion thus far [Papavasiliou et al., 2023].

The demand in zone D is distributed as in example 3.2, except for the distributions being uniform between 0 MWh and 400 MWh, and -400 MWh and 0 MWh. The combination of the probability distributions in zone B and zone D results in an equiprobable four-branch probability tree drawn from the random variables $\mathcal{U}[0, 100] + \mathcal{U}[0, 400]$ MWh, $\mathcal{U}[0, 100] + \mathcal{U}[-400, 0]$ MWh, $\mathcal{U}[-100, 0] + \mathcal{U}[0, 400]$ MWh, and $\mathcal{U}[-100, 0] + \mathcal{U}[-400, 0]$ MWh.

The equilibrium prices are presented in Table 3.6 and Fig. 3.5 compares the surplus distribution with respect to the "no adder" benchmark. The results are based on the aggregated offer curves that are generated from the optimal BSP bids in zone B and D. The aggregated curve is constructed with Eq. (3.18) and is described in the appendix. Consumer surplus refers to the cost of serving the inelastic BRP imbalance plus the capacity cost borne by all the consumers of zone B or D.

The platform prices from the "adder on BRPs and BSPs" and the "adder on BRPs" design are lower than the ones in the designs that induce truthful bidding, due to the lower offer curve in zone B. These altered offer curves result in an over-dispatch of the assets in zone B, and in an increased level of adders compared to the "RT market for reserve" design. The self-activation of assets for the "adder on BRPs" design generates particularly high imbalance prices in zone B.

Three adverse effects resulting from the "adder on BRPs and BSPs" and the "adder on BRPs" designs can be observed. First, the induced out-of-merit activations lead to an increased activation cost and an inefficient dispatch. The compliance with the objective outlined in article 3(m) of the Clean Energy Package [Commision, 2019] might be questioned.

Second, these designs give rise to cross-zonal distributive effects between consumers. The cost of decreasing the platform price is borne by the consumers in zone B, either through an increased imbalance price or through the capacity

Table 3.6: Example 3.3 – Expected prices (€/MWh)

	No adder	Adder on BRPs and BSPs	Adder on BRPs	RT market for reserve
Platform price	60.00	59.07	59.74	60.00
Scarcity component (zone B)	0.00	4.64	3.71	3.40
Balancing price (zone B)	60.00	63.71	63.45	63.40
Imbalance price (zone B)	60.00	63.71	67.08	63.40

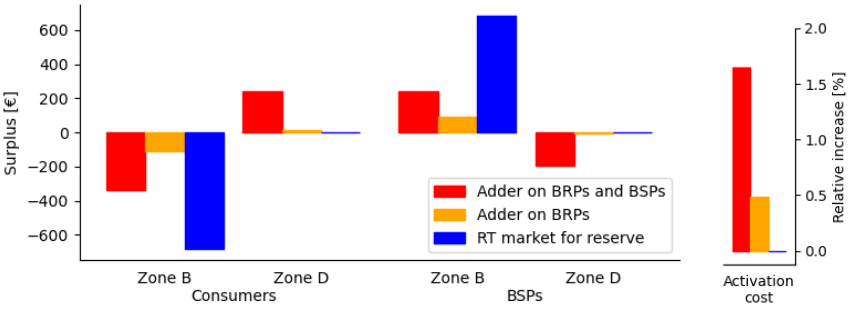


Figure 3.5: Differences in surplus and relative activation cost compared to the "no adder" benchmark in example 3.3.

cost. This suggests that the consumers in the zone with the adder subsidize the consumption of the consumers in the zone without the adder.

Third, these designs result in discrimination between BSPs from different zones. At similar marginal costs, BSPs in zone B are more likely to be activated than BSPs in zone D due to the increased balancing price or the possibility of resorting to reactive balancing. This leads to an increased surplus for BSPs in zone B compared to the "no-adder" benchmark and an opposite effect for zone D.

Only the "RT market for reserve" manages to introduce adders without inducing inefficiencies. In addition, this design only influences the surplus distribution between BRPs and BSPs in zone B and does not generate cross-zonal distributional effects.

The complete characterization of the equilibrium prices and surplus for each branch of the probability tree and for both zones can be found in the appendix.

3.5.4 Example 3.4 – Two Zones with Cross-Border Congestion and with Information on the Demand for Balancing Energy

We revisit example 3.3 by limiting the interconnector capacity between zone B and D at T MW. The analytical model with congestion requires the disaggregation of the demand for balancing energy between zone B and D. In a two-zone setting, congestion can be foreseen by BSPs in zone B when the difference between the demand for balancing energy in their zone and the activated balancing energy in their zone if there were no congestion, $x_B^{BSP,uncon}(x_B^{BRP}, x_D^{BRP})$, exceeds the interconnector capacity in either direction:

$$x_B^{BRP} - x_B^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}) \geq T, \tag{3.24}$$

or

$$x_B^{BRP} - x_B^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}) \leq -T. \tag{3.25}$$

The activated balancing energy in zone B as a function of the demand for balancing energy in zones B and D follows from these bounds and is presented in Eq. (3.26):

$$x_B^{BSP}(x_B^{BRP}, x_D^{BRP}) = \begin{cases} x_B^{BRP} + F \text{ if } x_B^{BRP} - x_B^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}) \geq T, \\ x_B^{BRP} - F \text{ if } x_B^{BRP} - x_B^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}) \leq -T, \\ x_B^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}) \text{ else.} \end{cases} \tag{3.26}$$

The platform price in zone B, $\lambda_{P,B}$, can be derived by evaluating the offer curve in zone B at the activated balancing energy,

$$\lambda_{P,B}(x_B^{BRP}, x_D^{BRP}) = B_B(x_B^{BSP}(x_B^{BRP}, x_D^{BRP})). \tag{3.27}$$

The same process can be reproduced to obtain the platform price and activated balancing energy in zone D.

Table 3.7 displays the expected zonal platform prices for an interconnector with a capacity of 50 MW. The expected prices are equal in both zones for the "no adder" and the "RT market for reserve" designs, due to the symmetrical bidding in upward and downward balancing energy.

Table 3.8 presents the relative increase in activation cost resulting from the suboptimal bidding for the "adder on BRPs" and the "adder on BRPs and BSPs" designs and for three levels of interconnector capacity: (i) the uncongested case with unlimited capacity (example 3.3), (ii) the case with some level of congestion (example 3.4), and (iii) the case with isolated systems (example 3.2). The last column reports the cost of the "no adder" design standardized to the case with unlimited capacity. Activation cost increases by up to 47%, depending on the availability of the interconnector.

Table 3.7: Example 3.4 – Expected prices (€/MWh) with an interconnector capacity of 50 MW in example 3.4.

	No adder	Adder on BRPs and BSPs	Adder on BRPs	RT market for reserve
Platform price (Zone B)	60.00	58.24	59.8	60.00
Platform price (Zone D)	60.00	59.54	59.9	60.00
Scarcity component	0.00	3.62	3.18	3.00

Table 3.8: Example 3.4 – Relative increase in activation cost (%) relative to least-cost activation

	Relative cost activation increase		Reference cost
Interconnector capacity	Adder on BRPs and BSPs	Adder on BRPs	No adder
$T = \infty$ (MW)	1.65	0.48	100.0
$T = 50$ (MW)	0.73	0.23	110.5
$T = 0$ (MW)	0.00	1.11 (5.56 for zone B)	147.1

The adverse effects encountered in example 3.3 are present for the case with some congestion, as displayed by the inefficient dispatch, but the efficiency losses due to limited cross-border capacity interfere with the analysis. There is also a question of how the congestion rent is allocated. Note that the relative increase in activation cost is not representative of the intensity of the distributional effects, as exemplified in figure 3.5 of the previous example. We have observed empirically that even minor changes in relative cost can be associated with significant changes in the distribution of welfare.

There is no distributional effect for the case with isolated systems. As already shown in example 3.1, the "adder on BRPs and BSPs" design does not generate inefficiencies in an isolated balancing market. The inefficiencies encountered in the "adder on BRPs" design is concentrated in zone B, the zone with the adder.

The complete characterization of the equilibrium prices and surplus for each branch of the probability tree and for both zones for an interconnector limit of 50 MW can be found in the appendix.

3.6 Conclusion

This paper investigates the unilateral application of an adder to balancing prices in a cross-border setting. An analytical model is used in order to identify the optimal strategies of flexibility providers under three different designs: the "adder on BRPs" design where the adder is applied to the imbalance price, the "adder on BRPs and BSPs" design, where the adder is applied to the balancing and imbalance price, and the "RT market for reserve" design that additionally introduces a real-time balancing capacity market. Market equilibria are derived based on these optimal strategies, extended to a cross-border setting, and illustrated on a two-zone example.

Adders, either on the imbalance price or on the balancing and imbalance price, in the absence of a real-time market for reserve, induce out-of-merit dispatch and increase the activation cost that is required for balancing the system. In a cross-border setting, this increased cost is borne by the consumers in the zone with the adder, as they face higher balancing and imbalance prices, whereas consumers in other zones enjoy lower prices. The introduction of a real-time market for reserve restores truthful bidding incentives and ensures that the increased cost to consumers in a zone, due to the adder, is fully distributed back to the flexibility suppliers in that zone.

3.A Optimal Strategies without Reservation Cost

The following strategies assume that the platform price, λ_P, obtained by clearing the balancing energy auction, is strictly monotone increasing in x, which is the level of required balancing activation. These strategies also assume that the supply function, λ_P^{-1}, is differentiable almost everywhere.[7] The scarcity component λ_R is assumed to be non-decreasing in the level of demand for balancing energy. A fringe agents can consider these two prices as being exogenous.

Proposition 3.8 (Bidding Strategy – No Adder)**.** *The optimal strategy for a fringe agent under a "no adder" design is to bid truthfully in the balancing auction.*

Proof. The optimal strategy for a fringe agent with upward balancing capacity P^+ and marginal cost C can be found by maximizing the sum of z_B and z_I.

a) If $\mathbb{E}_X[\lambda_P] \leq C$, then $z_I(q) = 0$ and the total reward of the agent $R(p,q)$ is defined hereunder.

$$R(p,q) = z_B(p,q) + z_I(q) \tag{3.28}$$
$$= C_1(p) \cdot q \tag{3.29}$$

with $C_1(p)$ characterized as follows:

$$C_1(p) = \int_{\lambda_P(x)>p} (\lambda_P(x) - C) dF(x) \tag{3.30}$$

Note that the bound on the integral in $C_1(p)$ can be reformulated, since λ_P is strictly monotonic increasing and thus has a uniquely defined inverse function.

$$\int_{\lambda_P(x)>p} (\lambda_P(x) - C) dF(x) = \int_{x>\lambda_P^{-1}(p)} (\lambda_P(x) - C) dF(x) \tag{3.31}$$

The optimal bidding strategy of the agent can be derived from the first-order conditions with respect to p and q. Let us find p first by fixing q.

$$\frac{\partial R(p,q)}{\partial p} = C_1'(p) \cdot q \tag{3.32}$$
$$= -(\lambda_P(\lambda_P^{-1}(p)) - C) \cdot f(\lambda_P^{-1}(p)) \cdot \frac{d\lambda_P^{-1}(p)}{dp} \cdot q \tag{3.33}$$

[7] At equilibrium, the supply function may exhibit breaking points due to BSPs resorting to reactive balancing. This can cause the supply function to not be differentiable everywhere.

$$= -(p - C) \cdot f(\lambda_P^{-1}(p)) \cdot \frac{d\lambda_P^{-1}(p)}{dp} \cdot q \tag{3.34}$$

For fixed q, the payoff function $R(p, q)$ is increasing in $(-\infty, C]$, zero at C, and decreasing in $[C, +\infty)$[8]. Thus for any q, an optimal strategy in the balancing auction is to bid truthfully the marginal cost. Given this strategy, the payoff becomes

$$R(C, q) = C_1(C) \cdot q. \tag{3.35}$$

We can then determine the first-order condition with respect to q.

$$\frac{\partial R(C, q)}{\partial q} = C_1(C) \tag{3.36}$$

$$= \int_{\lambda_P(x) > C} (\lambda_P(x) - C) dF(x) \tag{3.37}$$

$$> 0 \tag{3.38}$$

We conclude that, for a fringe agent with upward balancing capacity and marginal cost C higher than $\mathbb{E}_X[\lambda_P]$, the optimal strategy is to bid truthfully in the balancing auction.

b) If $\mathbb{E}_X[\lambda_P] > C$, then $z_I(q) = (\mathbb{E}_X[\lambda_P] - C) \cdot (P^+ - q)$ and the total reward of the agent $R(p, q)$ is defined hereunder.

$$R(p, q) = z_B(p, q) + z_I(q) \tag{3.39}$$

$$= C_1 + C_2 \cdot q + C_3(p) \cdot q \tag{3.40}$$

with C_1, C_2, and $C_3(p)$ characterized as follows:

$$C_1 = (\mathbb{E}_X[\lambda_P] - C) \cdot P^+ \tag{3.41}$$

$$C_2 = -(\mathbb{E}_X[\lambda_P] - C) \tag{3.42}$$

$$C_3(p) = \int_{\lambda_P(x) > p} (\lambda_P(x) - C) dF(x) \tag{3.43}$$

The optimal bidding strategy of the agent can be derived from the first-order conditions with respect to p and q. The optimal bidding price p can be obtained as in case a), and is equal to the marginal cost C.

Given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_2 \cdot q + C_3(C) \cdot q. \tag{3.44}$$

[8] This argument relies on $R(p, q)$ being continuous and $d\lambda_P^{-1}(p)/dp$ being positive where it exists, since λ_P is strictly monotonic increasing. The cases where $\lambda_P^{-1}(p)$ is not differentiable are analysed on a case-by-case basis.

We then have

$$\frac{\partial R(C,q)}{\partial q} = C_2 + C_3(C) \tag{3.45}$$

$$= -(\mathbb{E}_X[\lambda_P] - C) + \int_{\lambda_P(x)>C} (\lambda_P(x) - C)dF(x) \tag{3.46}$$

$$= -\left(\int_{x\leq\lambda_P^{-1}(C)} (\lambda_P(x) - C)dF(x) + \int_{x>\lambda_P^{-1}(C)} (\lambda_P(x) - C)dF(x) \right)$$

$$+ \int_{x>\lambda_P^{-1}(C)} (\lambda_P(x) - C)dF(x) \tag{3.47}$$

$$= -\int_{x\leq\lambda_P^{-1}(C)} (\lambda_P(x) - C)dF(x)) \tag{3.48}$$

$$> 0 \tag{3.49}$$

We conclude that, for a fringe agent with upward balancing capacity and marginal cost C higher than $\mathbb{E}_X[\lambda_P]$, the optimal strategy is to bid truthfully in the balancing auction.

□

Proposition 3.9 (Bidding Strategy – Adder on BRPs)**.** *The optimal strategy for a fringe agent under an "adder on BRPs" design is to bid truthfully in the balancing market if*

$$C \geq \mathbb{E}_X[\lambda_P + \lambda_R] - \int_{\lambda_P(x)\geq C} (\lambda_P(x) - C)dF(x),$$

else to perform reactive balancing.

Proof. The optimal strategy for a fringe agent with upward balancing capacity P^+ and marginal cost C can be found by maximizing the sum of z_B and z_I.

a) If $\mathbb{E}_X[\lambda_P + \lambda_R] \leq C$, then $z_I(q) = 0$ and the total reward of the agent $R(p,q)$ is defined hereunder.

$$R(p,q) = z_B(p,q) + z_I(q) \tag{3.50}$$

$$= C_1(p) \cdot q \tag{3.51}$$

with C_1 characterized as follows:

$$C_1 = \int_{\lambda_P(x)>p} (\lambda_P(x) - C)dF(x) \tag{3.52}$$

Note that the bound of the the integral in $C_1(p)$ can be reformulated as a function of x, since λ_P is strictly monotonic increasing and thus has an inverse function.

$$\int_{\lambda_P(x)>p} (\lambda_P(x) - C)dF(x) = \int_{x>\lambda_P^{-1}(p)} (\lambda_P(x) - C)dF(x) \tag{3.53}$$

The optimal bidding strategy of the agent can then be derived from the first-order conditions with respect to p and q. Let us focus on p first.

$$\frac{\partial R(p,q)}{\partial p} = C_1'(p) \cdot q \tag{3.54}$$

$$= -(\lambda_P(\lambda_P^{-1}(p)) - C) \cdot f(\lambda_P^{-1}(p)) \cdot \frac{d\lambda_P^{-1}(p)}{dp} \cdot q \tag{3.55}$$

$$= -(p - C) \cdot f(\lambda_P^{-1}(p)) \cdot \frac{d\lambda_P^{-1}(p)}{dp} \cdot q \tag{3.56}$$

For fixed q, the payoff function $R(p,q)$ is increasing in $(-\infty, C]$, zero at C, and decreasing in $[C, +\infty)$. Thus, for any q, an optimal strategy is to bid truthfully the marginal cost. Given this strategy, the payoff becomes

$$R(C,q) = C_1(C) \cdot q. \tag{3.57}$$

We can then determine the first order condition relative to q.

$$\frac{\partial R(C,q)}{\partial q} = C_1(C) \tag{3.58}$$

$$= \int_{\lambda_P(x)>C} (\lambda_P(x) - C)dF(x) \tag{3.59}$$

$$> 0 \tag{3.60}$$

We conclude that, for a fringe agent with upward balancing capacity and marginal cost C higher than $\mathbb{E}[\lambda_P + \lambda_R]$, the optimal strategy is to bid truthfully in the balancing auction.

b) If $\mathbb{E}[\lambda_P + \lambda_R] - \int_{\lambda_P(x)>C}(\lambda_P(x) - C)dF(x) \leq C < \mathbb{E}[\lambda_P + \lambda_R]$, then $z_I(q) = (\mathbb{E}[\lambda_P + \lambda_R] - C) \cdot (P^+ - q)$ and the total reward of the agent $R(p,q)$ is defined hereunder.

$$R(p,q) = z_B(p,q) + z_I(q) \tag{3.61}$$

$$= C_1 + C_2 \cdot q + C_3(p) \cdot q \tag{3.62}$$

with C_1, C_2 and $C_3(p)$ characterized as follows:

$$C_1 = (\mathbb{E}_X[\lambda_P + \lambda_R] - C) \cdot P^+ \tag{3.63}$$

$$C_2 = -(\mathbb{E}_X[\lambda_P + \lambda_R] - C) \tag{3.64}$$

$$C_3(p) = \int_{\lambda_P(x)>p} (\lambda_P(x) - C) dF(x) \tag{3.65}$$

The optimal bidding strategy of the agent can then be derived from the first-order conditions with respect to p and q. The optimal bidding price p can be obtained as in case a), and is equal to the marginal cost C.

Given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_2 \cdot q + C_3(C) \cdot q. \tag{3.66}$$

The next step is to examine the first-order condition with respect to q.

$$\frac{\partial R(C,q)}{\partial q} = C_2 + C_3(C) \tag{3.67}$$

$$= -(\mathbb{E}_X[\lambda_P + \lambda_R] - C) + \int_{\lambda_P(x)>C} (\lambda_P(x) - C) dF(x) \tag{3.68}$$

$$> 0 \tag{3.69}$$

We conclude that, for a fringe agent with upward balancing capacity P^+ and marginal cost C between $\mathbb{E}_X[\lambda_P + \lambda_R]$ and $\mathbb{E}_X[\lambda_P + \lambda_R] - \int_{\lambda_P(x)>C}(\lambda_P(x) - C)dF(x)$, the optimal strategy is to bid its entire capacity truthfully in the balancing auction.

c) If $C < \mathbb{E}_X[\lambda_P + \lambda_R] - \int_{\lambda_P(x)>C}(\lambda_P(x) - C)dF(x)$, then $z_I(q) = (\mathbb{E}_X[\lambda_P + \lambda_R] - C) \cdot (P^+ - q)$ and the total reward of the agent $R(p, q)$ is defined hereunder.

$$R(p, q) = z_B(p, q) + z_I(q) \tag{3.70}$$

$$= C_1 + C_2 \cdot q + C_3(p) \cdot q \tag{3.71}$$

with C_1, C_2 and $C_3(p)$ characterized as follows:

$$C_1 = (\mathbb{E}_X[\lambda_P + \lambda_R] - C) \cdot P^+ \tag{3.72}$$

$$C_2 = -(\mathbb{E}_X[\lambda_P + \lambda_R] - C) \tag{3.73}$$

$$C_3(p) = \int_{\lambda_P(x)>p} (\lambda_P(x) - C) dF(x) \tag{3.74}$$

The optimal bidding strategy of the agent can then be derived from the first-order conditions with respect to p and q. The optimal bidding price p can be obtained as in case a), and is equal to the marginal cost C.

Given this strategy, the payoff becomes

$$R(C, q) = C_1 + C_2 \cdot q + C_3(C) \cdot q. \tag{3.75}$$

The next step is to examine the first-order condition with respect to q.

$$\frac{\partial R(C,q)}{\partial q} = C_2 + C_3(C) \tag{3.76}$$

$$= -(\mathbb{E}[\lambda_P + \lambda_R] - C) + \int_{\lambda_P(x)>C} (\lambda_P(x) - C) dF(x) \tag{3.77}$$

$$< 0 \tag{3.78}$$

We conclude that, for a fringe agent with upward balancing capacity and marginal cost C lower than $\mathbb{E}_X[\lambda_P + \lambda_R] - \int_{\lambda_P(x)>C}(\lambda_P(x) - C)dF(x)$, the optimal strategy is to not participate in the balancing auction and to rather resort to reactive balancing.

□

Proposition 3.10 (Bidding Strategy – Adder on BRPs and BSPs)**.** *The optimal strategy for a fringe agent under an "adder on BRPs and BSPs" design is to bid its full capacity in the balancing energy market at price*

$$\lambda_P((\lambda_P + \lambda_R)^{-1}(C)).$$

Proof. The optimal strategy for a fringe agent with upward balancing capacity P^+ and marginal cost C can be found by maximizing the sum of z_B and z_I.

a) If $\mathbb{E}_X[\lambda_P + \lambda_R] \leq C$, then $z_I(q) = 0$ and the total reward of the agent $R(p,q)$ is defined hereunder.

$$R(p,q) = z_B(p,q) + z_I(q) \tag{3.79}$$

$$= C_1(p) \cdot q \tag{3.80}$$

with C_1 characterized as follows:

$$C_1 = \int_{\lambda_P(x)>p} (\lambda_P(x) + \lambda_R(x) - C) dF(x) \tag{3.81}$$

Note that the bound of the the integral in $C_3(p)$ can be reformulated as a function of x, since λ_P is strictly monotonic increasing and thus has an inverse function.

$$\int_{\lambda_P(x)>p} (\lambda_P(x)+\lambda_R(x)-C)dF(x) = \int_{x>\lambda_P^{-1}(p)} (\lambda_P(x)+\lambda_R(x)-C)dF(x) \tag{3.82}$$

The optimal bidding strategy of the agent can then be derived from the first-order conditions with respect to p and q. Let us find p first by fixing q.

$$\frac{\partial R(p,q)}{\partial p} = C_1'(p) \cdot q \tag{3.83}$$

$$= -(\lambda_P(\lambda_P^{-1}(p)) + \lambda_R(\lambda_P^{-1}(p)) - C) \cdot f(\lambda_P^{-1}(p)) \cdot \frac{d\lambda_P^{-1}(p)}{dp} \cdot q \tag{3.84}$$

$$= -(p + \lambda_R(\lambda_P^{-1}(p)) - C) \cdot f(\lambda_P^{-1}(p)) \cdot \frac{d\lambda_P^{-1}(p)}{dp} \cdot q \tag{3.85}$$

The unique root of $g(p) = p + \lambda_R(\lambda_P^{-1}(p)) - C$ is $\lambda_P(\lambda_I^{-1}(C))$, with $\lambda_I(x) = \lambda_P(x) + \lambda_R(x)$. This root can be interpreted as the platform price when the total system imbalance is such that the platform plus the reserve adder is equal to C.

$$g(\lambda_P(\lambda_I^{-1}(C))) = \lambda_P(\lambda_I^{-1}(C)) + \lambda_R(\lambda_P^{-1}(\lambda_P(\lambda_I^{-1}(C)))) - C \tag{3.86}$$

$$= \lambda_P(\lambda_I^{-1}(C)) + \lambda_R(\lambda_I^{-1}(C)) - C \tag{3.87}$$

$$= \lambda_I(\lambda_I^{-1}(C)) - C \tag{3.88}$$

$$= 0 \tag{3.89}$$

For fixed q, the payoff function $R(p, q)$ is increasing in $(-\infty, C']$, zero at C', and decreasing in $[C', +\infty)$ with $C' = \lambda_P(\lambda_I^{-1}(C))$. Thus, for any q, an optimal strategy in the balancing energy auction is to bid the platform price when the platform price plus the reserve adder is equal to the marginal cost of the agent. This strategy internalizes the added reward from the scarcity adder in the balancing bid such that the balancing bid is equal to the marginal cost minus the reserve adder when the platform price plus the reserve adder is equal to the marginal cost. Given this strategy, the payoff becomes

$$R(C', q) = C_1(C') \cdot q. \tag{3.90}$$

We can then determine the first order condition relative to q.

$$\frac{\partial R(C', q)}{\partial q} = C_1(C') \tag{3.91}$$

$$= \int_{\lambda_P(x) > C'} (\lambda_P(x) + \lambda_R(x) - C) dF(x) \tag{3.92}$$

$$= \int_{x > \lambda_I^{-1}(C)} (\lambda_I(x) - C) dF(x) \tag{3.93}$$

$$> 0 \tag{3.94}$$

We conclude that, for a fringe agent with upward balancing capacity and marginal cost C higher than $\mathbb{E}_X[\lambda_P + \lambda_R]$, the optimal strategy is to bid its full capacity at price $\lambda_P((\lambda_P + \lambda_R)^{-1}(C))$ in the balancing energy auction.

b) If $\mathbb{E}_X[\lambda_P + \lambda_R] > C$, then $z_I(q) = (\mathbb{E}_X[\lambda_P + \lambda_R] - C) \cdot (P^+ - q)$ and the total reward of the agent $R(p,q)$ is defined hereunder.

$$R(p,q) = z_B(p,q) + z_I(q) \tag{3.95}$$
$$= C_1 + C_2 \cdot q + C_3(p) \cdot q \tag{3.96}$$

with C_1 and $C_3(p)$ characterized as follows:

$$C_1 = (\mathbb{E}_{\mathbb{X}}[\lambda_P + \lambda_R] - C) \cdot P^+ \tag{3.97}$$
$$C_2 = -(\mathbb{E}_X[\lambda_P + \lambda_R] - C) \tag{3.98}$$
$$C_3(p) = \int_{\lambda_P(x) > p} (\lambda_P(x) + \lambda_R(x) - C) dF(x) \tag{3.99}$$

The optimal bidding strategy of the agent can then be derived from the first-order conditions with respect to p and q. The optimal bidding price p can be derived as previously, and is equal to $C' = \lambda_P(\lambda_I^{-1}(C))$.

Given this strategy, the payoff becomes

$$R(C', q) = C_1 + C_2 \cdot q + C_3(C') \cdot q. \tag{3.100}$$

We can then determine the first-order condition with respect to q.

$$\frac{\partial R(C', q)}{\partial q} = C_2 + C_3(C') \tag{3.101}$$
$$= -(\mathbb{E}_X[\lambda_P + \lambda_R] - C) + \int_{\lambda_P(x) > C'} (\lambda_P(x) + \lambda_R(x) - C) dF(x) \tag{3.102}$$
$$= -\left(\int_{x \leq \lambda_I^{-1}(C)} (\lambda_I(x) - C) dF(x) + \int_{x > \lambda_I^{-1}(C)} (\lambda_I(x) - C) dF(x) \right)$$
$$+ \int_{x > \lambda_I^{-1}(C)} (\lambda_I(x) - C) dF(x) \tag{3.103}$$
$$= -\int_{x \leq \lambda_I^{-1}(C)} (\lambda_I(x) - C dF(x)) \tag{3.104}$$
$$> 0 \tag{3.105}$$

We conclude that, for a fringe agent with upward balancing capacity and marginal cost C higher than $\mathbb{E}_X[\lambda_P + \lambda_R]$, the optimal strategy is to bid its entire capacity at price $\lambda_P(\lambda_I^{-1}(C))$ in the balancing energy auction.

□

Proposition 3.11 (Bidding Strategy – RT Market for Reserve)**.** *The optimal strategy for a fringe agent under a "RT market for reserve" design is to bid truthfully in the balancing energy market.*

Proof. The optimal strategy for a fringe agent with upward balancing capacity P^+ and marginal cost C can be found by maximizing the sum of z_B and z_I.

a) If $\mathbb{E}_X[\lambda_P] \leq C$, then $z_I(q) = 0$ and the total reward of the agent $R(p,q)$ is defined hereunder.

$$R(p,q) = z_B(p,q) + z_I(q) \tag{3.106}$$
$$= C_1 + C_2(p) \cdot q \tag{3.107}$$

with C_1 and $C_3(p)$ characterized as follows:

$$C_1 = \mathbb{E}_X[\lambda_R] \cdot P^+ \tag{3.108}$$
$$C_2(p) = \int_{\lambda_P(x)>p} (\lambda_P(x) - C) dF(x) \tag{3.109}$$

Maximizing $R(p,q)$ here is equivalent to case (a) of the "no adder" design, where we have shown that it is optimal to bid the entire capacity of the BSP truthfully.

b) If $\mathbb{E}_X[\lambda_P] > C$, then $z_I(q) = (\mathbb{E}_X[\lambda_P - C) \cdot (P^+ - q)$ and the total reward of the agent, $R(p,q)$, is defined hereunder.

$$R(p,q) = z_B(p,q) + z_I(q) \tag{3.110}$$
$$= C_1 + C_2 \cdot q + C_3(p) \cdot q \tag{3.111}$$

with C_1,C_2 and $C_3(p)$ characterized as follows:

$$C_1 = (\mathbb{E}_X[\lambda_P + \lambda_R] - C) \cdot P^+ \tag{3.112}$$
$$C_2 = -(\mathbb{E}_X[\lambda_P] - C) \tag{3.113}$$
$$C_3(p) = \int_{\lambda_P(x)>p} (\lambda_P(x) - C) dF(x) \tag{3.114}$$

Maximizing $R(p,q)$ here is equivalent to case (b) of the "no adder" design. The optimal strategy for an agent is to bid its entire capacity truthfully.

□

3.B Nash Equilibrium for "Adder on BRPs" Design

We define the platform price for a level α of reactive balancing,

$$\lambda_P(x,\alpha) = \begin{cases} MC(x-\alpha) & \text{if } x < \alpha, \\ \text{Price indeterminacy between } MC(0) \text{ and } MC(\alpha) & \text{if } x = \alpha, \\ MC(x) & \text{else.} \end{cases} \tag{3.115}$$

The opportunity cost of performing reactive balancing for an agent with marginal cost C, given a level α of reactive balancing in the system, is expressed as:

$$z(\alpha, C) = (\mathbb{E}_X[\lambda_P(\cdot,\alpha) + \lambda_R(\cdot)] - C) - \int_{\lambda_P(x,\alpha)\geq C} (\lambda_P(x,\alpha) - C) dF(x).$$

Proposition 3.12 (Equilibrium – Adder on BRPs)**.** *If $z(\alpha, MC(\alpha))$ is continuous, there exists a unique Nash equilibrium generated by fringe agents under the "adder on BRPs" design characterized by an equilibrium level of reactive balancing, α^*, such that $0 \leq \alpha^* \leq P^{max}$, and with the other BSPs bidding truthfully. This optimal level of reactive balancing is equal to (i) 0 if $z(0, MC(0)) < 0$, (ii) P^{max} if $z(P^{max}, MC(P^{max})) > 0$ or (iii) α^* characterized by the identity*

$$z(\alpha^*, MC(\alpha^*)) = 0. \tag{3.116}$$

This equilibrium level of reactive balancing generates platform prices equal to $\lambda_P(x, \alpha^*)$ and is strictly monotonic increasing.

Proof. $z(\alpha, C)$ is strictly monotonic decreasing in C for a fixed α as $z(\alpha, C) < z(\alpha, C^-)$ for $C^- < C$.

$$z(\alpha, C) - z(\alpha, C^-) = \int_{\lambda_P(x,\alpha)\leq C} (\lambda_P(x,\alpha)) - C) dF(x) - \int_{\lambda_P(x,\alpha)\leq C^-} (\lambda_P(x,\alpha) - C^-) dF(x) \tag{3.117}$$

$$= \int_{C^- \leq \lambda_P(x,\alpha)\leq C} (\lambda_P(x,\alpha) - C) dF(x) + \int_{\lambda_P(x,\alpha)\leq C^-} (C^- - C) dF(x) \tag{3.118}$$

$$< 0 \tag{3.119}$$

This allows us to prove the stability of the optimal level of reactive balancing mentioned in points (i) to (iii). As mentioned earlier, stability refers to a level of reactive balancing for which no agent has an incentive to deviate from their decision. BSPs after α^* on the merit order prefer participating in the balancing auction and agents before α^* prefer to resort to reactive balancing.

(i) If $z(0, MC(0)) < 0$, then a level of reactive balancing equal to 0 is stable as $z(0, MC(x)) < z(0, MC(0)$ for all x in $(0, P^{max}]$. In other words, if the cheapest generator finds the balancing energy auction more profitable than resorting to reactive balancing, every other generator should also find the balancing energy auction more profitable.

(ii) If $z(P^{max}, MC(P^{max})) > 0$, then a level of reactive balancing equal to P^{max} is stable as $z(P^{max}, MC(x)) > z(P^{max}, MC(x))$ for all x in $[0, P^{max})$. In other words, if the most expensive generator finds resorting to reactive balancing more profitable than participating in the balancing energy auction, every other generator should also find resorting to reactive balancing more profitable.

(iii) If $z(\alpha^*, MC(\alpha^*)) = 0$, then a level of reactive balancing equal to α^* is stable, since $z(\alpha^*, MC(\alpha^*)) > z(\alpha^*, MC(x))$ for all x in $(\alpha^*, P^{max}]$ and $z(\alpha^*, MC(\alpha^*)) < z(\alpha^*, MC(x))$ for all x in $[0, \alpha^*)$. In other words, if the frontier agent is indifferent between resorting to reactive balancing and participating in the balancing energy auction, every cheaper (resp. more expensive) generator should find reactive balancing (resp. the balancing energy auction) more profitable than the balancing energy auction (resp. doing reactive balancing) and has no incentive to modify its behaviour.

The existence of an equilibrium is proven by the continuity of $z(\alpha, MC(\alpha))$ with respect to α and the fact that at least one condition of (i) to (iii) must be true. The proof of uniqueness relies on $z(\alpha, MC(\alpha))$ being strictly monotonic decreasing with respect to α for strictly monotonic increasing MC. We show hereunder that $z(\alpha, MC(\alpha)) < z(\alpha^-, MC(\alpha^-))$ for $\alpha^- < \alpha$.

$$
\begin{aligned}
&z(\alpha^-, MC(\alpha^-)) - z(\alpha, MC(\alpha)) \\
&\qquad = \int_{x \le \alpha^-} (MC(x - \alpha^-) - MC(\alpha^-)) dF(x) \\
&\qquad\quad - \int_{x \le \alpha} (MC(x - \alpha) - MC(\alpha)) dF(x) \qquad (3.120) \\
&\qquad = \int_{x \le \alpha^-} (MC(x - \alpha^-) - MC(x - \alpha) - (MC(\alpha^-) - MC(\alpha))) dF(x) \\
&\qquad\quad - \int_{\alpha^- < x \le \alpha} MC(x - \alpha) - MC(\alpha) dF(x) \qquad (3.121) \\
&\qquad > 0 \qquad (3.122)
\end{aligned}
$$

This bidding behavior, coupled with the balancing energy auction selecting the bid in increasing price order, results in the platform price being equal to $MC(x - \alpha)$ for $x < \alpha$ (as $x - \alpha$ MWh of downward balancing capacity has to be activated) and to $MC(x)$ if $x > \alpha$.

The optimal strategy developed in proposition 2 of the main paper is still valid even when λ_P is not a standard single-valued function but rather a *set-valued function* with a price indeterminacy at α. The first-order condition with respect to p (3.54) exhibits two behaviours.

1. If $p < MC(0)$ and $p > MC(\alpha)$, the supply function λ_P^{-1} is well defined

and the first order condition is equal to the following:

$$\frac{\partial R(p,q)}{\partial p} = -(p-C)\cdot f(\lambda_P^{-1}(p))\cdot \frac{d\lambda_P^{-1}(p)}{dp}\cdot q. \tag{3.123}$$

2. If $MC(0) < p < MC(\alpha)$, the supply function hits a plateau (due to the agents between 0 and α resorting to reactive balancing and not participating in the energy balancing auction), and the payoff function is then characterized as follows.

$$R(p,q) = \int_{x>\alpha} (\lambda_P(x) - C)dF(x)\cdot q \tag{3.124}$$

The first order condition with respect to p is equal to 0 for any p between $MC(0)$ and $MC(\alpha)$.

The optimal price bid for an agent can be derived as a function of the agent's marginal cost, based on the continuity of R.

1. If $C < MC(0)$, the derivative of R with respect to p is increasing in $(-\infty, C[$, zero at C, decreasing in $]C, MC(0)[$, zero in $]MC(0), MC(\alpha)[$ and decreasing in $]MC(\alpha), \infty)$. Bidding $p = C$ is optimal.

2. If $MC(0) \leq C \leq MC(\alpha)$, the derivative of R with respect to p is increasing in $(-\infty, MC(0)[$, zero in $]MC(0), MC(\alpha)[$ and decreasing in $]MC(\alpha), \infty)$. Bidding any cost between $MC(0)$ and $MC(\alpha)$ (including $p = C$) is optimal.

3. If $MC(\alpha) < p$, the derivative of R with respect to p is increasing in $(-\infty, MC(0)[$, zero in $]MC(0), MC(\alpha)[$, increasing in $]MC(\alpha), C[$, zero at C and decreasing in $]C, \infty)$. Bidding $p = C$ is optimal.

□

The uniqueness of the equilibrium in the two-zone setting is not certain due to the impact of the level of reactive balancing on λ_R and its impact on the monotonicity of z.

The "adder on BRPs" design fails to produce a Nash equilibrium if the imbalance distribution is drawn from a unit set. In other words, if BSPs have perfect information on the level of imbalance that they will face, no equilibrium can be reached for the "adder on BRPs" design. The following argument illustrates this point. Let us assume that all agents know that the system will be exposed to a level of imbalance x'. In this context, the optimal level of reactive balancing is exactly x', since all agents below x' on the merit order would prefer the imbalance price to the balancing price and all agents after x' will participate in the balancing energy auction. The platform price could be anything between 0 and $MC(x')$ due to the discontinuity in the merit order curve.

a) If it is between 0 and $MC(x') - \lambda_R(x')$, a level x' of reactive balancing is not sustainable, as the agent at position x' on the merit order curve would rather participate in the balancing auction. But if the agent at position x' on the merit order curve participates truthfully in the balancing energy auction, they would be selected, and that platform price would have incentivized them to resort to reactive balancing.

b) If the platform price is between $MC(x') - \lambda_R(x')$ and $MC(x')$, then some agents located after x' on the merit order curve would rather do reactive balancing than participate in the balancing energy auction. The platform price with a level of reactive balancing higher than x' is not sufficient to sustain this level of self-activation.

Only a balancing energy price equal to $MC(x') - \lambda_R(x')$ would ensure a Nash equilibrium in this situation but, even if it is possible, it is not guaranteed.

Note that there is no risk of failing to produce a Nash equilibrium even if the domain of the imbalance distribution is a unit set in the 2-zones setting. The aggregation of a continuous and a non-continuous offer curve is continuous and this induces unique balancing energy prices.

3.C Optimal Strategies with Reservation Cost

The following strategies are derived for profit-maximizing agents with production cost C and reservation cost K. These agents can first participate in a reservation auction followed by a balancing energy auction for activation. They can also only participate in the balancing energy auction or only resort to reactive balancing. The reservation cost is incurred once if (i) the agent is selected in the reservation auction, and/or (ii) if the agent is selected in the balancing energy auction, and/or (iii) if the agent resorts to reactive balancing. The agents offer a reservation bid (p^{DA}, q^{DA}), a balancing energy bid (p, q), and decide to self-activate ai MWh. The reservation payoff can be expressed as:

$$z_{DA}(p^{DA}, q^{DA}) = \begin{cases} \lambda_{DA} \cdot q^{DA} & \text{if } \lambda_{DA} \geq p^{DA}, \\ 0 & \text{else.} \end{cases} \tag{3.125}$$

Here, λ_{DA} is the reservation price. If the reservation bid is accepted, the balancing energy payoff is:

$$z_B(p, q, q^{DA}, x) = \begin{cases} \lambda_P \cdot (q + q^{DA}) & \text{if } \lambda_P(x) \geq p, \\ 0 & \text{else,} \end{cases} \tag{3.126}$$

and the expected balancing energy payoff is

$$z_B(p, q, q^{DA}) = \int_{\lambda_P(x) \geq p} (\lambda_P(x) - C) dF(x) \cdot (q + q^{DA}). \tag{3.127}$$

Finally, the reactive balancing payoff if the reservation bid has been accepted is the solution to the following optimization problem:

$$\max_{ai} \quad (\mathbb{E}_X[\lambda_{imb}] - C) \cdot ai \tag{3.128}$$

$$s.t. \quad ai + q + q^{DA} \leq P^+ \tag{3.129}$$

$$ai \geq 0 \tag{3.130}$$

If $C \geq \mathbb{E}_X[\lambda_{imb}]$, the optimal level of reactive balancing $ai^\star$ is 0, else $ai^\star$ is equal to the leftover capacity from the balancing energy and reservation auctions. The reactive balancing payoff is then described as follows:

$$z_I(q, q^{DA}) = \begin{cases} (\mathbb{E}_X[\lambda_{imb}] - C) \cdot (P^+ - q - q^{DA}) & \text{if } C \leq \mathbb{E}_X[\lambda_{imb}], \\ 0 & \text{else.} \end{cases} \tag{3.131}$$

The introduction of a real-time market for reserve has an impact on the real-time payoff. The capacity that was accepted in the reservation auction has to be bought back in real-time:

$$\begin{aligned} & z_B(p, q, q^{DA}, x) + z_I(ai, x) \\ &= \begin{cases} (\lambda_P(x) + \lambda_R(x) - C) \cdot (q + q^{DA} + ai) + \lambda_R(x) \cdot (P^+ - q - q^{DA} - ai) - \lambda_R(x) \cdot q^{DA} \\ \qquad \text{if } \lambda_P(x) \leq p, \\ (\lambda_P(x) + \lambda_R(x) - C) \cdot ai + \lambda_R(x) \cdot (P^+ - ai) - \lambda_R(x) \cdot q^{DA} \\ \qquad \text{else,} \end{cases} \end{aligned} \tag{3.132}$$

$$= \begin{cases} (\lambda_P(x) - C) \cdot (q + q^{DA} + ai) + \lambda_R(x) \cdot (P^+ - q^{DA}) & \text{if } \lambda_P(x) \leq p, \\ (\lambda_P(x) - C) \cdot ai + \lambda_R(x) \cdot (P^+ - q^{DA}) & \text{else.} \end{cases} \tag{3.133}$$

Both the expected imbalance and balancing payoffs can be modified to account for the real-time market for reserve:

$$z_I(q, q^{DA}) = \begin{cases} (\mathbb{E}_X[\lambda_P] - C) \cdot (P^+ - q - q^{DA}) & \text{if } C \leq \mathbb{E}_X[\lambda_P] \\ 0 & \text{else,} \end{cases} \tag{3.134}$$

$$z_B(p, q, ^{DA}) = \int_{\lambda_P(x) \geq p} (\lambda_P(x) - C) dF(x) \cdot (q + q^{DA}) + \mathbb{E}_X[\lambda_R] \cdot (P^+ - q^{DA}). \tag{3.135}$$

We can now derive the optimal bidding strategy for a fringe agent with marginal cost C and reservation cost K. The reservation cost can be considered as a fixed must-run cost for keeping the plant up-and-running.

Proposition 3.13 (Bidding Strategy – Reservation Cost – No Adder). *The optimal strategy for a fringe agent under a "no adder" design is to bid its full capacity in the day-ahead auction at a price*

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; 0\right),$$

with $\delta z_B = \int_{\lambda_P(x)>C}(\lambda_P(x) - C)dF(x)$*, and to then bid truthfully in the balancing auction. If the day-ahead bid is not accepted, the optimal action is to not activate its plant.*

Proof. (a) If $E[\lambda_P] \leq C$ then $z_I(q, q^{DA}) = 0$. The total reward of an agent selected in the day-ahead auction, $R(p, q, q^{DA})$, is defined hereunder for a given day-ahead price λ^{DA}:

$$R(p, q, q^{DA}) = z_B(p, q) + z_I(q) + z_{DA}(q^{DA}) \tag{3.136}$$
$$= C_1 + C_2(p) \cdot q + C_3 \cdot q^{DA} + C_4(p) \cdot q^{DA} \tag{3.137}$$

with $C_1, C_2(p), C_3$ and $C_4(p)$ equal to:

$$C_1 = -K \tag{3.138}$$
$$C_2(p) = \int_{\lambda_P(x)>p}(\lambda_P(x) - C)dF(x) \tag{3.139}$$
$$C_3 = \lambda^{DA} \tag{3.140}$$
$$C_4(p) = \int_{\lambda_P(x)>p}(\lambda_P(x) - C)dF(x) \tag{3.141}$$

The optimal bidding strategy of the agent can be derived from the first-order condition with respect to p, q and q^{DA}. We focus on p first:

$$\frac{\partial R(p, q, q^{DA})}{\partial p} = C_2'(p) \cdot q + C_4'(p) \cdot q^{DA} \tag{3.142}$$
$$= -(\lambda_P(\lambda_P^{-1}(p)) - C) \cdot f(\lambda_P^{-1}(p)) \cdot \frac{d\lambda_P^{-1}(p)}{dp} \cdot (q + q^{DA}) \tag{3.143}$$
$$= -(p - C) \cdot f(\lambda_P^{-1}(p)) \cdot \frac{d\lambda_P^{-1}(p)}{dp} \cdot (q + q^{DA}) \tag{3.144}$$

For fixed q and q^{DA}, the payoff function $R(p, q, q^{DA})$ is increasing in $(-\infty, C]$, zero at C, and decreasing in $[C, +\infty)$. Thus, for any q and q^{DA}, an optimal strategy in the balancing auction is to bid truthfully the marginal cost.

The payoff maximization problem given this strategy is presented in equation (3.145). It shows that, for a positive λ^{DA}, the coefficient on q^{DA} in the objective is higher than on q. The strategy of setting q at 0 dominates all other strategies for allocating the balancing capacity between q and q^{DA}.

$$\max_{q, q^{DA}} \quad R(C, q, q^{DA}) \tag{3.145}$$

$$\begin{aligned} s.t. \quad & q^{DA} + q \leq P^+ \\ & q^{DA} \geq 0 \\ & q \geq 0 \end{aligned}$$

Given these two strategies, the payoff becomes

$$R(C, 0, q^{DA}) = C_1 + C_3 \cdot q^{DA} + C_4(C) \cdot q^{DA}. \tag{3.146}$$

The first-order condition with respect to q^{DA} can be examined, in order to compute the optimal action with respet to q^{DA}.

$$\begin{aligned} \frac{\partial R(C, 0, q^{DA})}{\partial q^{DA}} &= C_2 + C_4(C) && (3.147) \\ &= \lambda^{DA} + \int_{\lambda_P(x) > C'} (\lambda_P(x) - C) dF(x) && (3.148) \\ &= \lambda^{DA} + \int_{x > \lambda_P^{-1}(C)} (\lambda_P(x) - C) dF(x) && (3.149) \\ &> 0 && (3.150) \end{aligned}$$

We conclude that, for a fringe agent with upward balancing capacity and marginal cost C higher than $\mathbb{E}[\lambda_P]$ selected in the day-ahead auction at an arbitrary day-ahead price λ^{DA}, the optimal strategy is to bid its full capacity in the DA auction and to then bid thruthfully in the balancing auction.

The payoff of an agent selected in the day ahead given this strategy is expressed as a function of a given day-ahead price as follows:

$$R^{DA}(\lambda_{DA}) = \delta z_B \cdot P^+ - K + \lambda_{DA} \cdot P^+. \tag{3.151}$$

The payoff for an agent not participating in the day-ahead auction can be inferred from proposition 3.8:

$$R^B = \begin{cases} \delta z_B \cdot P^+ - K & \text{if } K \leq \Pi_B, \\ 0 & \text{else.} \end{cases} \tag{3.152}$$

The optimal price bid in the day ahead can then be derived from the profit maximization problem (3.153).

$$Q(p^{DA}) = \max_{p^{DA}} \int_{\lambda_{DA}(x) \leq p^{DA}} R^B dF^{DA}(x) + \int_{\lambda_{DA}(x) > p^{DA}} R^{DA}(\lambda_{DA}(x)) dF^{DA}(x) \tag{3.153}$$

This problem is defined as a function of the stochastic day-ahead demand x, the corresponding day-ahead price $\lambda^{DA}(x)$ and the CDF and PDF of the demand $F^{DA}(\cdot)$ and $f^{DA}(\cdot)$. The total payoff is the probability-weighted sum of (i) the payoff of an agent not selected in the day-ahead auction, if $p^{DA} \geq \lambda_{DA}$, and (ii) the payoff of an agent selected in the day ahead, if $p^{DA} < \lambda_{DA}$.

If the balancing payoff is lower than the reservation cost, so $\Pi_B < K$ and $R^B = 0$, the optimal bidding strategy can be derived from the first order condition with respect to p^{DA}:

$$\frac{dQ(p^{DA})}{dp^{DA}} = -R^{DA}(\lambda_{DA}(\lambda_{DA}^{-1}(p^{DA}))) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.154}$$

$$= -(\delta z_B \cdot P^+ - K + p^{DA} \cdot P^+) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA}) \cdot \frac{d\lambda_D A^{-1}(p^{DA})}{dp^{DA}} \tag{3.155}$$

Condition (3.155) indicates that the optimal price bid in the day ahead is $p^{DA} = (\delta z_B \cdot P^+ - K)/P^+$.

The same reasoning can be applied if the balancing payoff is greater than the reservation cost, so $\delta z_B \cdot P^+ > K$ and $R^B = \delta z_B \cdot P^+ - K$:

$$\frac{\partial Q(p^{DA})}{\partial p^{DA}} = (\delta z_B - K - R^{DA}(p^{DA})) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.156}$$

$$= -p^{DA} \cdot P^+ \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.157}$$

Condition (3.157) specifies that the optimal price bid in the day ahead is $p^{DA} = 0$.

The optimal bidding strategy in the day-ahead market if $\mathbb{E}[\lambda_P] \leq C$ is then for a unit to bid its maximal capacity at price

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; 0\right)$$

and to then bid truthfully in the balancing energy auction, if selected, or not activate its plant, if not selected.

(b) If $E[\lambda_P] > C$, then $z_I(q, q^{DA}) = \mathbb{E}[\lambda_P - C] \cdot (P^+ - q - q^{DA})$. Since participating in the balancing energy auction is always more profitable than resorting to reactive balancing, the reasoning developed in point (a) applies and the optimal bidding strategy in the day-ahead market is to bid the full capacity at price

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; 0\right)$$

and to then bid truthfully in the balancing energy auction, if selected, or not activate the plant, if not selected. □

Proposition 3.14 (Bidding Strategy – Reservation Cost – Adder on BRPs)**.** *The optimal strategy for a fringe agent under an "adder on BRPs" design is to bid its full capacity in the day-ahead auction at price*

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; \delta z_B - \delta z_I; 0\right)$$

with $\delta z_B = \int_{\lambda_P(x)>C}(\lambda_P(x)-C)dF(x)$ *and* $\delta z_I = (\mathbb{E}[\lambda_P+\lambda_R]-C)$. *If the DA bid has been accepted, the optimal action is to bid truthfully in the balancing auction. If the DA bid is* $\delta z_I - \delta z_B$ *and has not been accepted, the optimal action for an agent is to do reactive balancing with its full capacity. Else, the agent should not activate its plant.*

Proof. (a) If $\mathbb{E}[\lambda_P + \lambda_R] \leq C$, then $z_I(q, q^{DA}) = 0$. As participating in the balancing energy auction is always more profitable than resorting to reactive balancing, the reasoning developed in point (a) of proposition 3.13 applies and the optimal bidding strategy in the day-ahead market is to bid the maximal capacity at a price

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; 0\right)$$

and to then bid truthfully in the balancing energy auction, if selected, or not activate the plant, if not selected.

(b) If $E[\lambda_P + \lambda_R] - \int_{\lambda_P(x)>C}(\lambda_P(x) - C)dF(x) \leq C < \mathbb{E}[\lambda_P + \lambda_R]$, then $z_I(q, q^{DA}) = \mathbb{E}[\lambda_P + \lambda_R - C] \cdot (P^+ - q - q^{DA})$. Since participating in the balancing energy auction is always more profitable than resorting to reactive balancing, the reasoning developed in point (a) of proposition 3.13 applies and the optimal bidding strategy in the day-ahead market is to bid the maximal capacity at a price

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; 0\right)$$

and to then bid truthfully in the balancing energy auction, if selected, or not activate the plant, if not selected.

(c) If $C < E[\lambda_P + \lambda_R] - \int_{\lambda_P(x)>C}(\lambda_P(x) - C)dF(x)$ then $z_I(q, q^{DA}) = \mathbb{E}[\lambda_P + \lambda_R - C] \cdot (P^+ - q - q^{DA})$ and the total reward of an agent selected in the day-ahead auction, $R(p, q, q^{DA})$, is defined hereunder for an arbitrary day-ahead price λ^{DA}:

$$R(p, q, q^{DA}) = z_B(p, q) + z_I(q) + z_{DA}(q^{DA}) \tag{3.158}$$

$$= C_1 + C_2 \cdot q + C_3(p) \cdot q + C_4 \cdot q^{DA} + C_5(p) \cdot q^{DA} \tag{3.159}$$

with $C_1, C_2, C_3(p), C_4$ and $C_5(p)$ expressed as:

$$C_1 = -K + (\mathbb{E}[\lambda_P + \lambda_R] - C) \cdot P^+ \tag{3.160}$$

$$C_2 = -(\mathbb{E}[\lambda_P + \lambda_R] - C) \tag{3.161}$$

$$C_3(p) = \int_{\lambda_P(x)>p}(\lambda_P(x) - C)dF(x) \tag{3.162}$$

$$C_4 = -(\mathbb{E}[\lambda_P + \lambda_R] - C) + \lambda_{DA} \tag{3.163}$$

$$C_5(p) = \int_{\lambda_P(x)>p}(\lambda_P(x) - C)dF(x) \tag{3.164}$$

The same argument as in case (a) of proposition 3.13 can be applied for proving the optimality of bidding $p = C$ in the balancing auction. Similarly, the strategy consisting of bidding its full capacity in the reservation auction dominates all other strategies for allocating the capacity between the reservation auction and the balancing energy auction.

The first order condition with respect to q^{DA} can be examined next:

$$\frac{\partial R(C,0,q^{DA})}{\partial q^{DA}} = C_4 + C_5(C) \tag{3.165}$$

$$= -(\mathbb{E}[\lambda_P + \lambda_R] - C) + \lambda^{DA} + \int_{\lambda_P(x)>C} (\lambda_P(x) - C) dF(x) \tag{3.166}$$

$$= \lambda_{DA} + \delta z_B - \delta z_I \tag{3.167}$$

In contrast to case (a) and (b), it is not possible to ascertain the sign of $\partial R(C,O,q^{DA})\partial q^{DA}$ for any positive λ^{DA}. If $\lambda^{DA} > \delta z_I - \delta z_B$ then

$$\frac{\partial R(C,0,q^{DA})}{\partial q^{DA}} > 0 \quad \text{and} \quad q^{DA} = P^+$$

and if $\lambda^{DA} < \delta z_I - \delta z_B$ then

$$\frac{\partial R(C,0,q^{DA})}{\partial q^{DA}} < 0 \quad \text{and} \quad q^{DA} = 0.$$

The payoff for an agent selected in the day-ahead market is then dependent on the value of λ^{DA} and on the day-ahead capacity bid q^{DA}. It is characterized as follows, as a function of an arbitrary day-ahead price:

$$R^{DA}(\lambda_{DA}, q^{DA}) = q^{DA} \cdot (\lambda_{DA} + \delta z_B) + (P^+ - q^{DA}) \cdot \delta z_I - K \tag{3.168}$$

The payoff for an agent not participating in the day-ahead auction can be derived from proposition 3.9:

$$R^B = \begin{cases} \delta z_I \cdot P^+ - K & \text{if } K \leq \delta z_I \cdot P^+, \\ 0 & \text{else.} \end{cases} \tag{3.169}$$

The optimal price and quantity bid in the day ahead can then be derived from the profit maximization problem (3.170).

$$Q(p^{DA}, q^{DA}) = \max_{q^{DA}, p^{DA}} \int_{\lambda_{DA}(x) \leq p^{DA}} R^B dF^{DA}(x) + \int_{\lambda_{DA}(x) > p^{DA}} R^{DA}(\lambda_{DA}(x), q^{DA}) dF^{DA}(x) \tag{3.170}$$

The total payoff is the probability weighted sum of (i) the payoff of an agent not selected in the day-ahead auction, if $p^{DA} \geq \lambda_{DA}$, and (ii) the payoff of an agent selected in the day ahead, if $p^{DA} < \lambda_{DA}$:

If $\delta z_I \cdot P^+ < K$ then $R^B = 0$, and the optimal strategy can be derived from the first order condition with respect to p^{DA} and q^{DA}:

$$\frac{\partial Q(p^{DA}, q^{DA})}{\partial p^{DA}} = -R^{DA}(\lambda_{DA}(\lambda_{DA}^{-1}(p^{DA})), q^{DA}) \cdot F^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.171}$$

$$= -R^{DA}(p^{DA}, q^{DA}) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.172}$$

$$= (P^+ - q^{DA}) \cdot \delta z_I - K) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}}$$
$$- (q^{DA} \cdot (p^{DA} + \delta z_B) \tag{3.173}$$

Equating (3.173) to zero leads to the following optimality condition:

$$p^{DA*} = \frac{K - P^+ \cdot \delta z_I + q^{DA} \cdot (\delta z_I - \delta z_B)}{q^{DA}} \tag{3.174}$$

This allows us to define the optimal price bid p^{DA*} for any capacity bid q^{DA} different from 0. We can also derive the first-order condition with respect to q^{DA}:

$$\frac{\partial Q(p^{DA}, q^{DA})}{\partial q^{DA}} = \int_{\lambda_{DA}(x) > p^{DA}} (\lambda_{DA}(x) + \delta z_B - \delta z_I) dF^{DA}(x). \tag{3.175}$$

The next step consists in inserting p^{DA*} in equation (3.175):

$$\frac{\partial Q(p^{DA*}, q^{DA})}{\partial q^{DA}} = \int_{\lambda_{DA}(x) > p^{DA*}} (\lambda_{DA}(x) + \delta z_B - \delta z_I) dF^{DA}(x) \tag{3.176}$$

$$> 0 \tag{3.177}$$

The last inequality is due to p^{DA*} always being greater than $\delta z_B - \delta z_I$ for any q^{DA} different than zero. This leads to $q^{DA*} = P^+$ and $p^{DA*} = (K - P^+ \cdot \delta z_B)/P^+$.

If $\delta z_I \cdot P^+ \geq K$ then $R^B = \delta z_I \cdot P^+ - K$ and the optimal strategy can be derived from the first-order condition with respect to p^{DA} and q^{DA}:

$$\frac{\partial Q(p^{DA}, q^{DA})}{\partial p^{DA}} = (R^B - R^{DA}(p^{DA}, q^{DA})) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.178}$$

$$= (\delta z_I \cdot P^+ - K - q^{DA} \cdot (p^{DA} + \delta z_B) - (P^+ - q^{DA}) \cdot \delta z_I + K)$$
$$\cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.179}$$

$$= -(p^{DA} + \delta z_B - \delta z_I) \cdot q^{DA} \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.180}$$

For fixed q^{DA}, the day-ahead payoff function $R^{DA}(p^{DA}, q^{DA})$ is increasing in $(-\infty, \delta z_I - \delta z_B[$, zero at $\delta z_I - \delta z_B$ and decreasing in $]\delta z_I - \delta z_B, +\infty)$. Thus, for any q^{DA}, an optimal strategy in the reservation auction is to bid $\delta z_I - \delta z_B$.

We can also derive the first-order condition with respect to q^{DA}:

$$\frac{\partial Q(\delta z_I - \delta z_B, q^{DA})}{\partial q^{DA}} = \int_{\lambda_{DA}(x) > \delta z_I - \delta z_B} (\lambda_{DA}(x) + \delta z_B - \delta z_I) dF^{DA}(x) \tag{3.181}$$

$$\geq 0 \tag{3.182}$$

This shows that the optimal strategy in the day-ahead capacity auction is to bid the full capacity at the price $\delta z_I - \delta z_B$.

Combining both cases $\delta z_I \cdot P^+ < K$ and $\delta z_I \cdot P^+ \geq K$ allows us to define the optimal bid in the capacity auction if $C < \mathbb{E}[\lambda_P - \lambda_R] - \int_{\lambda_P(x) \geq C} (\lambda_P(x) - C) dF(x)$ as

$$p^{DA} = \max\left(\delta z_I - \delta z_B, \frac{K - \delta z_B \cdot P^+}{P^+}\right). \tag{3.183}$$

□

Proposition 3.15 (Bidding Strategy – Reservation Cost – Adder on BRPs and BSPs)**.** *The optimal strategy for a fringe agent under an "adder on BRPs and BSPs" design is to bid its full capacity in the day-ahead auction at price*

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; 0\right),$$

with $\delta z_B = \int_{\lambda_P(x) > \lambda_P((\lambda_P + \lambda_R)^{-1}(C))} (\lambda_P(x) - C) dF(x)$*, and to then bid at price* $\lambda_P((\lambda_P + \lambda_R)^{-1}(C))$ *in the balancing energy auction. If the day-ahead bid is not accepted, the optimal action is to not activate its plant.*

Proof. (a) If $E[\lambda_P] \leq C$, then $z_I(q, q^{DA}) = 0$. As internalizing the value of the adder when participating in the balancing energy auction is always more profitable than resorting to reactive balancing, the reasoning developed in point (a) in proposition 3.16 applies and the optimal bidding strategy in the day-ahead market is to bid the maximal capacity at price

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; 0\right)$$

and to then internalize the value of the adder in the balancing energy auction, if selected, or to not activate the plant, if not selected.

(b) If $E[\lambda_P] > C$, then $z_I(q, q^{DA}) = \mathbb{E}[\lambda_P + \lambda_R - C] \cdot (P^+ - q - q^{DA})$. Since internalizing the value of the adder when participating in the balancing energy auction is always more profitable than resorting to reactive balancing,

the reasoning developed in point (a) in proposition 3.16 applies, and the optimal bidding strategy in the day-ahead market is to bid the full capacity at price

$$p^{DA} = \max\left(\frac{K - \delta z_B \cdot P^+}{P^+}; 0\right)$$

and to then internalize the value of the adder in the balancing energy auction, if selected, or not activate the plant, if not selected. □

Proposition 3.16 (Bidding Strategy – Reservation Cost – RT Market for Reserve)**.** *The optimal strategy for a fringe agent under a "RT market for reserve" design is to bid its full capacity in the day-ahead auction at price*

$$p^{DA} = \max\left(\frac{K - (\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+}{P^+}; \mathbb{E}_X[\lambda_R]\right),$$

with $\delta z_B = \int_{\lambda_P(x)>C}(\lambda_P(x) - C)dF(x)$ *and* $\delta z_I = \mathbb{E}[\lambda_P] - C$*. If the day-ahead bid is accepted, the optimal action is to bid truthfully in the balancing auction. If the day-ahead bid is* $\mathbb{E}_X[\lambda_R]$ *and has not been accepted, the optimal action for the agent is to bid truthfully in the balancing auction with its full capacity. Else, the agent should not activate its plant.*

Proof. (a) If $\mathbb{E}[\lambda_P] \leq C$ then $z_I(q, q^{DA}) = 0$ and the total reward of an agent selected in the day-ahead auction, $R(p, q, q^{DA})$, is defined hereunder for a given day-ahead price λ^{DA}:

$$R(p, q, q^{DA}) = z_B(p, q) + z_I(q) + z_{DA}(q^{DA}) \tag{3.184}$$

$$= C_1 + C_2(p) \cdot q + C_3 \cdot q^{DA} + C_4(p) \cdot q^{DA} \tag{3.185}$$

with $C_1, C_2(p), C_3$ and $C_4(p)$ equal to:

$$C_1 = -K + \mathbb{E}_X[\lambda_R] \cdot P^+ \tag{3.186}$$

$$C_2(p) = \int_{\lambda_P(x)>p} (\lambda_P(x) - C)dX(x) \tag{3.187}$$

$$C_3 = -\mathbb{E}[\lambda_R] + \lambda_{DA} \tag{3.188}$$

$$C_4(p) = \int_{\lambda_P(x)>p} (\lambda_P(x) - C)dX(x) \tag{3.189}$$

The same argument as in case (a) of proposition 3.13 can be applied for proving the optimality of bidding $p = C$ in the balancing auction. The dominance of allocating the capacity of an agent between q and q^{DA} depends on the sign of $\lambda_{DA} - \mathbb{E}_X[\lambda_R]$.

If $\lambda_{DA} > \mathbb{E}_X[\lambda_R]$, then the payoff of the real-time reserve market is less profitable than the day-ahead payoff, and setting q at 0 dominates all other strategies for allocating the balancing capacity between q and q^{DA}. If $\lambda_{DA} < \mathbb{E}_X[\lambda_R]$, then the real-time reserve payoff is more profitable than the day-ahead

payoff and setting q^{DA} at 0 dominates all other strategies for allocating the balancing capacity between q and q^{DA}. In either case, resorting to reactive balancing is dominated by at least one of these strategies.

The payoff for an agent selected in the day ahead is then dependent on the value of λ^{DA} and on the day-ahead capacity bid q^{DA}. It is expressed as follows, as a function of a given day-ahead price:

$$R^{DA}(\lambda_{DA}, q^{DA}) = (\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ + (\lambda_{DA} - \mathbb{E}_X[\lambda_R]) \cdot q^{DA} - K \quad (3.190)$$

The payoff for an agent not participating in the day-ahead auction can be derived from proposition 3.9:

$$R^B = \begin{cases} (\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ - K & \text{if } K \leq (\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+, \\ 0 & \text{else.} \end{cases} \quad (3.191)$$

The optimal price and quantity bid in the day-ahead market can then be derived from the profit maximization problem (3.192).

$$Q(p^{DA}, q^{DA}) = \max_{q^{DA}, p^{DA}} \int_{\lambda_{DA}(x) \leq p^{DA}} R^B dF^{DA}(x) + \int_{\lambda_{DA}(x) > p^{DA}} R^{DA}(\lambda_{DA}(x), q^{DA}) dF^{DA}(x) \quad (3.192)$$

The total payoff is the probability weighted sum of (i) the payoff of an agent not selected in the day-ahead auction, if $p^{DA} \geq \lambda_{DA}$, and (ii) the payoff of an agent selected in the day ahead, if $p^{DA} < \lambda_{DA}$.

If $(\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ < K$ then $R^B = 0$, and the optimal strategy can be derived from the first-order condition with respect to p^{DA} and q^{DA}:

$$\frac{\partial Q(p^{DA}, q^{DA})}{\partial p^{DA}} = -R^{DA}(\lambda_{DA}(\lambda_{DA}^{-1}(p^{DA})), q^{DA}) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \quad (3.193)$$

$$= -R^{DA}(p^{DA}, q^{DA}) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \quad (3.194)$$

$$= (p^{DA} - \mathbb{E}_X[\lambda_R]) \cdot q^{DA} - K) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}}$$
$$- ((\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ \quad (3.195)$$

Equating (3.173) to zero leads to the following optimality condition:

$$p^{DA*} = \frac{K - P^+ \cdot (\delta z_B + \mathbb{E}_X[\lambda_R])}{q^{DA}} + \mathbb{E}_X[\lambda_R] \quad (3.196)$$

This allows us to define the optimal price bid p^{DA*} for any capacity bid q^{DA} different from 0. We can also derive the first-order condition with respect to q^{DA}:

$$\frac{\partial Q(p^{DA}, q^{DA})}{\partial q^{DA}} = \int_{\lambda_{DA}(x) > p^{DA}} (\lambda_{DA}(x) - \mathbb{E}_X[\lambda_R]) dF^{DA}(x). \quad (3.197)$$

The next step consists of inserting p^{DA*} in equation (3.197):

$$\frac{\partial Q(p^{DA*}, q^{DA})}{\partial q^{DA}} = \int_{\lambda_{DA}(x) > p^{DA*}} (\lambda_{DA}(x) - \mathbb{E}_X[\lambda_R]) dF^{DA}(x) \tag{3.198}$$

$$> 0 \tag{3.199}$$

The last inequality is due to p^{DA*} always being greater than or equal to $\mathbb{E}_X[\lambda_R]$ for $K > P^+ \cdot (\delta z_B + \mathbb{E}_X[\lambda_R])$ for any q^{DA} different from zero and positive. This leads to $q^{DA*} = P^+$ and

$$p^{DA*} = (K - P^+ \cdot \delta z_B)/P^+.$$

If $(\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ \geq K$ then $R^B = (\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ - K$ and the optimal strategy can be derived from the first-order condition with respect to p^{DA} and q^{DA}:

$$\frac{\partial Q(p^{DA}, q^{DA})}{\partial p^{DA}} = (R^B - R^{DA}(p^{DA}, q^{DA})) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.200}$$

$$= (\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ - K - ((\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ + (p^{DA} - \mathbb{E}_X[\lambda_R]) \cdot q^{DA} - K)) \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.201}$$

$$= -(p^{DA} - \mathbb{E}_X[\lambda_R]) \cdot q^{DA} \cdot f^{DA}(\lambda_{DA}^{-1}(p^{DA})) \cdot \frac{d\lambda_{DA}^{-1}(p^{DA})}{dp^{DA}} \tag{3.202}$$

For fixed q^{DA}, the day-ahead payoff function $R^{DA}(p^{DA}, q^{DA})$ is increasing in $(-\infty, \mathbb{E}_X[\lambda_R][$, zero at $\mathbb{E}_X[\lambda_R]$, and decreasing in $]\mathbb{E}_X[\lambda_R], +\infty)$. Thus, for any q^{DA}, an optimal strategy in the reservation auction is to bid $\mathbb{E}_X[\lambda_R]$.

We can also derive the first-order condition with respect to q^{DA}:

$$\frac{\partial Q(\mathbb{E}_X[\lambda_R], q^{DA})}{\partial q^{DA}} = \int_{\lambda_{DA}(x) > \mathbb{E}_X[\lambda_R]} (\lambda_{DA}(x) - \mathbb{E}_X[\lambda_R]) dF^{DA}(x) \tag{3.203}$$

$$> 0 \tag{3.204}$$

This shows that the optimal strategy in the day-ahead capacity auction is to bid the entire capacity at the price $\mathbb{E}_X[\lambda_R]$.

Combining both cases $(\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ < K$ and $(\delta z_B + \mathbb{E}_X[\lambda_R]) \cdot P^+ \geq K$ allows us to define the optimal bid in the capacity auction if $C < \mathbb{E}_X[\lambda_P]$ as

$$p^{DA} = \max\left(\mathbb{E}_X[\lambda_R], \frac{K - \delta z_B \cdot P^+}{P^+}\right). \tag{3.205}$$

(b) If $E[\lambda_P] > C$ then $z_I(q, q^{DA}) = \mathbb{E}[\lambda_P - C] \cdot (P^+ - q - q^{DA})$. Since participating in the balancing energy auction is always more profitable than

resorting to reactive balancing, the reasoning developed in point (a) applies and the optimal bidding strategy in the day-ahead market is to bid the maximal capacity at price

$$p^{DA} = \max\left(\mathbb{E}_X[\lambda_R], \frac{K - \delta z_B \cdot P^+}{P^+}\right).$$

If the reservation bid is accepted, the optimal action is to bid truthfully in the balancing auction. If the day-ahead bid is $\mathbb{E}_X[\lambda_R]$ and is not accepted, the optimal action is to bid the full capacity truthfully in the balancing energy auction. Else, the agent should not activate its plant. □

3.D Aggregated Offer Curves in a Cross-Border Setting with Linear Marginal Cost

This section describes the analytical formulation characterizing the market equilibrium resulting from the aggregation of two zones: B and D. Zone B has a merit order curve equal to $MC_B(x) = a_B x + b$ and applies one of the four designs with an ORDC equal to $ORDC(x) = a_R x$. Zone D has a merit order curve equal to $MC_D(x) = a_D x + b$.

3.D.1 No Adder and RT Market for Reserve:

The aggregated offer curve is defined as follows:

$$B(x) = B_B(x) \cup B_D(x) = MC_B(x) \cup MC_D(x). \tag{3.206}$$

This results in

$$\begin{cases} a_D x_D = a_B x_B \\ x_D + x_B = x \end{cases} \implies \begin{cases} x_B = \frac{a_D x}{a_D + a_B} \\ x_D = \frac{a_B x}{a_D + a_B} \end{cases} \tag{3.207}$$

and

$$B(x) = MC_B\left(\frac{x a_D}{a_D + a_B}\right) = \frac{a_D a_B x}{a_D + a_B} + b. \tag{3.208}$$

3.D.2 Adder on BRPs and BSPs:

Under this design, $B_B(x) = MC_B(x) - \lambda_R(x)$ so a_B is replaced by $a_B - a_R$ if $x \geq 0$.

$$B(x) = \begin{cases} \frac{a_D a_B x}{a_D + a_B} + b & \text{if } x \leq 0, \\ \frac{a_D (a_B - a_R) x}{a_D + a_B - a_R} + b & \text{else.} \end{cases} \tag{3.209}$$

3.D.3 Adder on BRPs:

Under this design,

$$B_B(x) = \begin{cases} MC_B(x-\alpha) & \text{if } x \leq \alpha, \\ MC_B(x) & \text{else.} \end{cases} \tag{3.210}$$

Due to the discontinuity in $B_B(x)$, three cases have to be considered: (i) $x \leq \alpha$, $\alpha < x \leq (1 + a_B/a_D)\alpha$, and (iii) $x > (1 + a_B/a_D)\alpha$. The results are presented here are for an arbitrary α.

$$x_B = \begin{cases} \frac{a_D(x-\alpha)}{a_D+a_B} & \text{if } x \leq \alpha \\ 0 & \text{if } \alpha < x \leq (1 + a_B/a_D)\alpha \\ \frac{a_D x}{a_B+a_D} - \alpha & \text{if } x > (1 + a_B/a_D)\alpha \end{cases} \tag{3.211}$$

$$x_D = \begin{cases} \frac{a_B(x-\alpha)}{a_D+a_B} & \text{if } x \leq \alpha \\ x - \alpha & \text{if } \alpha < x \leq (1 + a_B/a_D)\alpha \\ \frac{a_B x}{a_B+a_D} & \text{if } x > (1 + a_B/a_D)\alpha \end{cases} \tag{3.212}$$

$$B(x) = \begin{cases} \frac{a_B a_D(x-\alpha)}{a_D+a_B} + b & \text{if } x \leq \alpha \\ a_D(x-\alpha) + b & \text{if } \alpha < x \leq (1 + a_B/a_D)\alpha \\ \frac{a_B a_D x}{a_B+a_D} + b & \text{if } x > (1 + a_B/a_D)\alpha \end{cases} \tag{3.213}$$

3.E Analytical Platform Prices with Congestion

This section revisits the analytical formulation of the market equilibrium resulting from the connection of zones B and D with an interconnector capacity of T MW. As previously, the merit order curves in zone B and D are $MC_B(x) = a_B x + b$ and $MC_D(x) = a_D x + b$. The ORDC in zone B is equal to $ORDC(x) = a_R x$. The demand for balancing energy is denoted as x^{BRP} and the activated balancing energy as x^{BSP}. The platform price in zone B and D is denoted as $\lambda_{P,B}$ and $\lambda_{P,D}$ respectively.

3.E.1 No Adder and RT Market for Reserve

The offer curve in zone B and D is equal to the merit order curve. If there is no congestion,

$$x_B^{BSP,uncon} = K_D \cdot (x_B^{BRP} + x_D^{BRP}) \tag{3.214}$$

$$x_D^{BSP,uncon} = K_B \cdot (x_B^{BRP} + x_D^{BRP}). \tag{3.215}$$

with $K_B = a_B/(a_B + a_D)$ and $K_D = a_D/(a_B + a_D)$. The interconnector is congested when

$$T \leq x_B^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}) - x_B^{BRP} \text{ or } T \leq x_B^{BRP} - x_B^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}). \tag{3.216}$$

Note that, in a two-zone setting, this is equivalent to

$$T \leq x_D^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}) - x_D^{BRP} \text{ or } T \leq x_D^{BRP} - x_D^{BSP,uncon}(x_B^{BRP}, x_D^{BRP}). \tag{3.217}$$

These constraints allow us to compute the activated balancing energy in both zones for all cases: no congestion, congestion B to D, and congestion D to B.

$$x_B^{BSP} = \begin{cases} x_B^{BRP} + T & \text{if } T \leq K_D \cdot (x_B^{BRP} + x_D^{BRP}) - x_B^{BRP} \\ x_B^{BRP} - T & \text{if } T \leq x_B^{BRP} - K_D \cdot (x_B^{BRP} + x_D^{BRP}) \\ K_D \cdot (x_B^{BRP} + x_D^{BRP}) & \text{else} \end{cases} \tag{3.218}$$

$$x_D^{BSP} = \begin{cases} x_D^{BRP} - T & \text{if } T \leq K_D \cdot (x_B^{BRP} + x_D^{BRP}) - x_B^{BRP} \\ x_D^{BRP} + T & \text{if } T \leq x_B^{BRP} - K_D \cdot (x_B^{BRP} + x_D^{BRP}) \\ K_B \cdot (x_B^{BRP} + x_D^{BRP}) & \text{else.} \end{cases} \tag{3.219}$$

The platform price in both zones can then be expressed as follows:

$$\lambda_{P,B}(x_B^{BRP}, x_D^{BRP}) = B_B(x_B^{BSP}) \tag{3.220}$$

$$= \begin{cases} MC_B(x_B^{BRP} + T) & \text{if } T \leq K_D \cdot (x_B^{BRP} + x_D^{BRP}) - x_B^{BRP} \\ MC_B(x_B^{BRP} - T) & \text{if } T \leq x_B^{BRP} - K_D \cdot (x_B^{BRP} + x_D^{BRP}) \\ MC_B(K_D \cdot (x_B^{BRP} + x_D^{BRP})) & \text{else} \end{cases} \tag{3.221}$$

$$\lambda_{P,D}(x_B^{BRP}, x_D^{BRP}) = B_D(x_D^{BSP}) \tag{3.222}$$

$$= \begin{cases} MC_D(x_D^{BRP} - T) & \text{if } T \leq K_D \cdot (x_B^{BRP} + x_D^{BRP}) - x_B^{BRP} \\ MC_D(x_D^{BRP} + T) & \text{if } T \leq x_B^{BRP} - K_B \cdot (x_B^{BRP} + x_D^{BRP}) \\ MC_D(K_D \cdot (x_B^{BRP} + x_D^{BRP})) & \text{else} \end{cases} \tag{3.223}$$

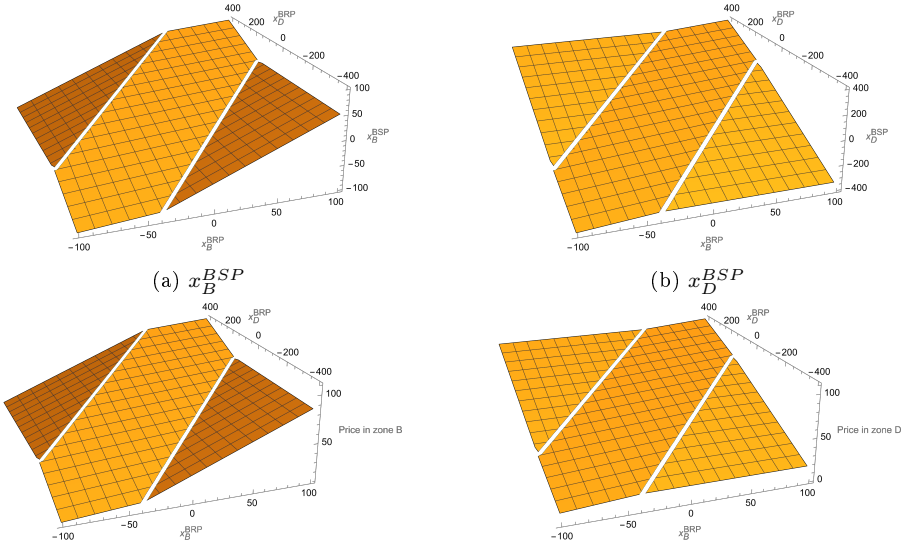
This equilibrium is illustrated in figure 3.6.

3.E.2 Adder on BRPs and BSPs

If there is no congestion, the offer curves in zone B and D are respectively equal to the merit order curve in zone B minus the ORDC and the merit order curve in zone D. The activated balancing energy can then be expressed as follows:

$$x_B^{BSP} = \begin{cases} K_D \cdot (x_B^{BRP} + x_D^{BRP}) & \text{if } x_B^{BRP} + x_D^{BRP} \leq 0 \\ K_D^R \cdot (x_B^{BRP} + x_D^{BRP}) & \text{else} \end{cases} \tag{3.224}$$

$$x_D^{BSP} = \begin{cases} K_B \cdot (x_B^{BRP} + x_D^{BRP}) & \text{if } x_B^{BRP} + x_D^{BRP} \leq 0 \\ K_B^R \cdot (x_B^{BRP} + x_D^{BRP}) & \text{else.} \end{cases} \tag{3.225}$$



(a) x_B^{BSP} (b) x_D^{BSP}

(c) $\lambda_{P,B}$ (d) $\lambda_{P,D}$

Figure 3.6: Illustration of the activated balancing energy and the price at equilibrium under the "no adder" and "RT market for reserve" designs for $a_B = 1/2$, $a_D = 1/8$, $b = 60$ and $T = 50$

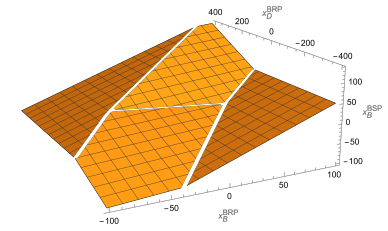
Here, $K_B^R = (a_B - a_R)/(a_D + a_B - a_R)$ and $K_D^R = a_D/(a_B + a_D - a_R)$. From there, the same procedure as for the "no adder" design can be reproduced to generate six cases: no congestion, congestion from D to B and congestion from B to D for either negative or positive demand for balancing energy. The equilibrium is illustrated in figure 3.7.

3.E.3 Adder on BRPs

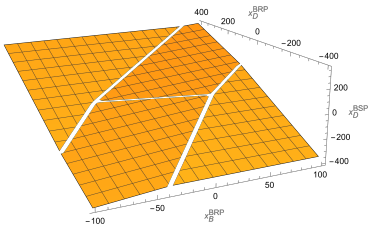
If there is no congestion, the offer curves in zone B and D can be expressed as follows:

$$x_B^{BSP} = \begin{cases} K_D \cdot (x_B^{BRP} + x_D^{BRP} - \alpha) & \text{if } x_B^{BRP} + x_D^{BRP} \leq \alpha \\ 0 & \text{if } \alpha \leq x_B^{BRP} + x^{BRP_D} \leq (1 + a_B/a_D)\alpha \\ K_D \cdot (x_B^{BRP} + x_D^{BRP}) - \alpha & \text{else} \end{cases} \tag{3.226}$$

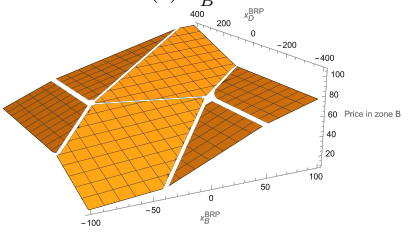
$$x_D^{BSP} = \begin{cases} K_B \cdot (x_B^{BRP} + x_D^{BRP} - \alpha) & \text{if } x_B^{BRP} + x_D^{BRP} \leq \alpha \\ x_B^{BRP} + x_D^{BRP} - \alpha & \text{if } \alpha \leq x_B^{BRP} + x^{BRP_D} \leq (1 + a_B/a_D)\alpha \\ K_B \cdot (x_B^{BRP} + x_D^{BRP}) & \text{else.} \end{cases} \tag{3.227}$$



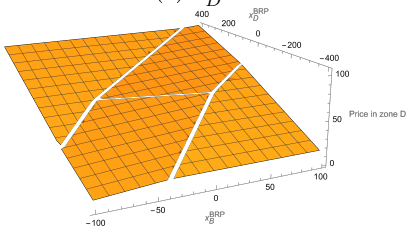
(a) x_B^{BSP}



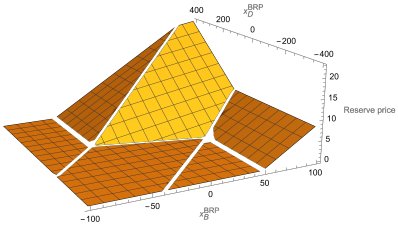
(b) x_D^{BSP}



(c) $\lambda_{P,B}$



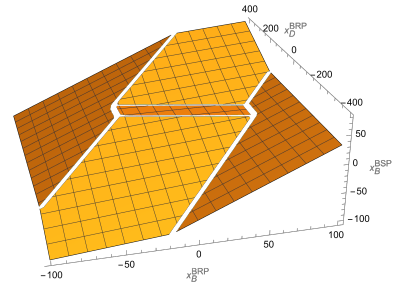
(d) $\lambda_{P,D}$



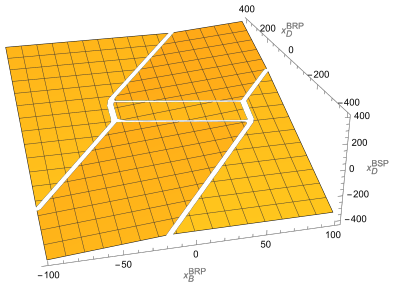
(e) λ_R

Figure 3.7: Illustration of the activated balancing energy and prices at equilibrium under the "adder on BRPs and BSPs" design for $a_B = 1/2$, $a_D = 1/8$, $a_R = 1/6$, $b = 60$ and $T = 50$

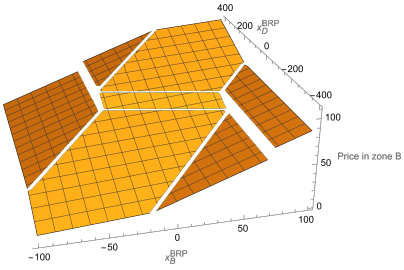
The same procedure as previously can be reproduced to generate nine cases: no congestion, congestion from D to B and congestion from B to D for either a demand for balancing energy lower than α, a demand for balancing energy between α and $(1 + a_B/a_D)\alpha$ and a demand for balancing energy greater than $(1 + a_B/a_D)\alpha$. Figure 3.8 illustrates these results. Note that, for the non-congested case when $\alpha \leq x_B^{BRP} + x_D^{BRP} \leq (1 + a_B/a_D)\alpha$, there is no activated balancing energy from zone B and we need to use the offer curve and the activated balancing energy in zone D to obtain the platform price.



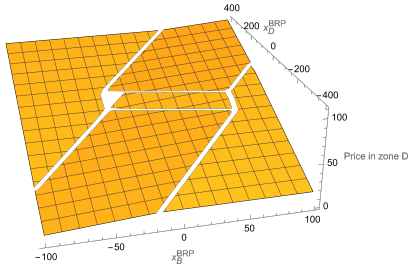
(a) x_B^{BSP}



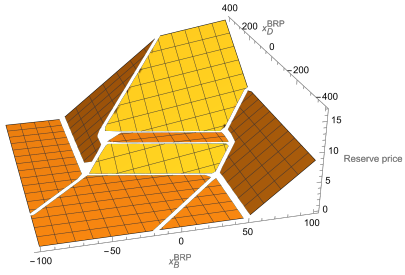
(b) x_D^{BSP}



(c) $\lambda_{P,B}$



(d) $\lambda_{P,D}$



(e) λ_R

Figure 3.8: Illustration of the activated balancing energy and prices at equilibrium under the "adder on BRPs" design for $a_B = 1/2$, $a_D = 1/8$, $a_R = 1/6$, $\alpha = 20$, $b = 60$ and $T = 50$

4 Multi-Product Balancing Markets

4.1 Introduction

A second source of difference between the balancing and imbalance prices is generated by the multiplicity of balancing energy product. Balancing the market is a continuous process and, broadly speaking, it involves the activation of fast balancing products with high reactivity over a short duration versus slow balancing products with low reactivity over a longer duration. This tradeoff is not unique to European systems. Every electricity system has to deal with electricity imbalances to ensure the stability of the power grid. The fast balancing product considered in this chapter is the *automatic frequency restoration reserve* (aFRR) product of the European market, which can react almost instantaneously to a dispatch order, whereas the slow balancing product analyzed in this work is the *manual frequency restoration reserve* (mFRR) product of the European market. The balancing process is composed of a sequence of short-term energy auctions for aFRR preceded by one longer-term energy auction for mFRR. aFRR balancing energy is priced every four seconds and mFRR balancing energy every fifteen minutes. Both balancing energy products are priced independently. Similarly to mFRR, imbalance is priced once every fifteen minutes. The *imbalance settlement harmonization methodology* (ISHM) provides guidelines for forming the imbalance price based on the aFRR and mFRR platform prices and this chapter proposes a game-theoretical framework for analysing these guidelines and understanding the interaction between different balancing products in a multi-product balancing system comprised of multiple distinct balancing energy auctions. Two issues are specifically investigated in this work: (i) the strategy of the system operator for activating balancing energy, and (ii) the imbalance settlement scheme.

The activation strategy of the system operator refers to the trade-off that is faced by the system operator when allocating the imbalance between balancing products. Balancing the market is a continuous process and, broadly speaking, it involves the activation of fast balancing products with high reactivity over a short duration versus slow balancing products with low reactivity over longer

duration. This tradeoff is not unique to European systems. Every electricity system has to deal with electricity imbalances to ensure the stability of the power grid. The fast balancing product studied in this paper is the *automatic frequency restoration reserve* (aFRR) product of the European market, which can react almost instantaneously to a dispatch order and has a full activation time of 5 minutes. In contrast, the slow balancing product analyzed in this work is the *manual frequency restoration reserve* (mFRR) product of the European market. We assume that it cannot be adjusted in the very short term. The balancing process can be modelled as a sequence of short-term energy auctions for aFRR preceded by one longer-term energy auction for mFRR. The standard practice in Europe is a sequence of 225 four-second aFRR auctions preceded by a fifteen-minute mFRR auction. In terms of operation, the system operator sets a demand for mFRR balancing energy over the next fifteen minutes and the leftover imbalance, the original imbalance minus the mFRR balancing energy, is covered every four seconds in the aFRR activation auctions. The system operator activation strategy refers to the process for determining the demand for mFRR balancing energy. Reasoning about this process as a two-stage decision process, the activation strategy can be interpreted as a first-stage decision whereas the demand for aFRR balancing energy is a recourse decision given the first-stage decision. The activation strategy of the system operator will be referred to as the *mFRR activation strategy* for the remainder of the paper.

The imbalance pricing scheme refers to the mechanism for pricing the demand side of balancing markets, which is represented by price-inelastic imbalances. These imbalances are generated by agents that are connected to the grid and deviate from their traded positions due to forecast errors in renewable energy supply or electricity demand, the sudden loss of assets due to failure, or the intentional deviations from market schedules for supporting the system or seizing a better price to maximimize profit, to name a few reasons.[1] Imbalances are settled over fifteen-minute intervals, that are referred to as *imbalance settlement periods* (ISP), at the *imbalance price*. TSOs have some freedom concerning the design of the imbalance price, however there exists a decision by the European Union Agency for the Cooperation of Energy Regulators, ACER, to harmonize and standardize imbalance settlement [ACER, 2020b]. This decision is referred to as the *imbalance settlement harmonization methodology*, ISHM, wherein article 9 states that the imbalance price should be based on the mFRR and aFRR prices. This paper reviews three imbalance pricing schemes considered in the power market design discourse: (i) the "mean mFRR and aFRR" price, (ii) the "max mFRR and mean aFRR" prices, and (iii) the "mFRR only" price. The first two are based on the ISHM and constructed by respectively taking the mean of the mFRR and aFRR prices and the maximum of the mFRR and the mean aFRR price, and the "mFRR only" price accounts solely for the mFRR price when computing the imbalance settlement price.

Both elements (the TSO activation strategy and the imbalance pricing

[1] Agents generating imbalances are referred to as *balance responsible parties* (BRPs) in European balancing markets.

scheme) affect the payoff that is associated with the actions of agents and their bidding behavior. The mFRR activation strategy impacts the expected aFRR and mFRR payoffs of the agents. The imbalance pricing scheme modifies the payoff of the self-dispatching option for flexible assets. Certain imbalance pricing schemes can incentivize flexible assets to opt out of the centralized aFRR and mFRR auctions and to self-activate their asset in order to be intentionally in imbalance and be remunerated at the imbalance price.

The investigation in this paper is related to ongoing revisions of imbalance pricing schemes throughout Europe such as the ratification of the ISHM in the Nordic countries (Norway, Sweden, Finland and Denmark) [Stattnett et al., 2023]. The discussion in these countries revolves around shifting from the "mFRR only" imbalance pricing scheme to the "mean mFRR and aFRR" or the "max mFRR and mean aFRR" approaches.

An additional motivation for this work is connected to the price incidents that occurred in Austria following its connection to the balancing platforms MARI and PICASSO. Austria has experienced aFRR prices above 7500 €/MWh for 0.17% of the PICASSO platform optimization runs over the last 6 months of 2022 [ACER, 2023]. This corresponds to 7.5 hours of extreme prices, which impact the Austrian imbalance cost. The study of the mFRR activation strategy was recommended by ACER as a means for mitigating the aFRR price incidents.

Additionally to the balancing energy auctions and imbalance settlement, the balancing process also includes *balancing capacity auctions*. They are held in the day ahead, before or after the day-ahead wholesale energy market, in order to ensure an adequate supply of real-time balancing energy. In contrast to balancing energy auctions, balancing capacity auctions have remained at the national level. Every TSO has specific procedures for procuring the capacity. A common practice is to rely on "pay-as-bid" for pricing but discussions at the European level about cross-border exchange of balancing capacity[2] and the co-optimization of wholesale energy and balancing capacity in an integrated day-ahead market [Papavasiliou and Avila, 2024] indicate a progressive move to a "pay-as-clear" (i.e. uniform pricing) mechanism. This analysis can be extended to include balancing capacity market but this feature is ignored for the remainder of the paper.

Two European idiosyncracies that are not included in this analysis are *dual-pricing* schemes for imbalance settlement and *direct mFRR activation*. The ISHM permits Member States to employ a dual-pricing structure. In this structure, the settlement of imbalances for a market participant is contingent upon the direction of its own imbalance in relation to the aggregate system position. This investigation considers a *single-pricing* scheme with a unique imbalance settlement price for all agents generating imbalances. Direct mFRR activation is the process that allows system operators to submit a demand for mFRR

[2]See ACER's website `https://www.acer.europa.eu/electricity/market-rules/electricity-balancing/capacity-calculation-and-allocation` and [Cho and Papavasiliou, 2024] for an application.

balancing energy during the course (as opposed to the beginning of) an imbalance settlement period. These direct activations are typically triggered by generation contingencies. Only *scheduled mFRR activations* occurring at the beginning of an ISP are accounted for in our model.

Our analysis aims at contributing to the electricity market design literature. Analytical and game-theoretical techniques have been used in [Fabra et al., 2006] to compare pay-as-bid and uniform pricing in energy auctions under conditions of market power. Similar frameworks are also used in [Bushnell and Oren, 1994] and [Chao and Wilson, 2002] in order to analyse coupled reserve and energy auctions. [Bushnell and Oren, 1994] analyze scoring rules with discriminatory pricing in the reserve auction and [Chao and Wilson, 2002] prove, through backward induction, that independent capacity and energy auctions induce a truthful revelation of cost under uniform pricing and in the presence of price-taking agents. Multi-product capacity auctions are investigated in [Kamat and Oren, 2002] without an energy component. Similar methods have been applied in the analysis of European balancing markets. [Ocker et al., 2018] focus on pricing rules for balancing markets and strategic interactions between agents. [Ehrhart and Ocker, 2021] include the day-ahead wholesale market in their analysis. [Cartuyvels et al., 2023] examine the uncoordinated implementation of adders in integrated energy auctions and introduce an outside option to balancing energy auctions through imbalance settlement.

Non-analytical methods have also been used in order to address specific questions of market design in European balancing markets. Agent-based models have been used for investigating the effect of the imbalance pricing scheme [van der Veen et al., 2012], market organization [Poplavskaya et al., 2020], the introduction of free bids [Poplavskaya et al., 2021], and the back-propagation of real-time balancing capacity prices to day-ahead markets [Papavasiliou et al., 2021]. [Petitet et al., 2019] use a simulation-based model in order to analyze the impact of the gate-closure time on operating cost.

Our paper extends the literature on balancing market design by proposing an analysis for the case of multiple reserve products that accounts for real-time balancing constraints and the intricate relationship between faster- and slower-moving reserves. The goal of this paper is to provide a quantitative framework for highlighting the incentives that are generated in a multi-product balancing market and the sensitivity of the market equilibrium to (i) the imbalance pricing scheme, and (ii) the activation strategy for slow-moving reserve (mFRR in our case).

The four main policy insights uncovered by our analysis are summarized as follows. (1) The "mFRR only" imbalance price allows for simple optimal strategies and prevents self-dispatching from participants. (2) Minimum balancing activation cost can be reached from the "mFRR only" and "mean mFRR and aFRR" imbalance pricing schemes under the optimal mFRR activation strategy. (3) If the optimal mFRR activation strategy is not available, the "mean mFRR and aFRR" imbalance pricing scheme incentivizes self-dispatching in a way that reduces the balancing cost compared to the no-reaction benchmark.

(4) The "max mFRR and mean aFRR" imbalance price distorts price signals and induces an inefficient level of self-dispatching. This translates to actionable presprictive policy guidance for the ongoing European market design debates that are mentioned previously in the introduction.

The remainder of the paper is structured as follows. Section 4.2 presents the multi-product balancing market model that is used in our analysis. Section 4.3 analyzes the impact of the mFRR activation strategy and the imbalance pricing scheme on the balancing energy equilibrium. Section 4.4 illustrates the results of our analytical models on an example and section 4.6 discusses them. Section 4.7 concludes.

4.2 Modelling Multi-Product Balancing Markets

This section describes in detail the different components of our model. It begins by describing the social optimum for a benevolent system operator with direct control over flexible assets. The second subsection describes the mechanism used in practice for emulating this resource allocation (the sequence of balancing energy markets). The individual strategy of fringe agents participating in the game is described. The last subsection discusses the aggregation of those strategy sets and their mapping to aFRR and mFRR merit orders.

4.2.1 Least-Cost Activation

The social optimum for covering the imbalances is attained when a system operator minimizes the activation cost of flexible assets while respecting balancing constraints. System operators have two technologies at their disposal for covering imbalances: slow capacity that is dispatched once at the beginning of the horizon (a 15-minute imbalance settlement period in our analysis) and cannot be modified later and fast capacity that can modify its dispatch as uncertainty unfolds. The least-cost activation problem can be modeled as a two-stage stochastic program (4.1) and (4.2):[3]

$$\min_{x^S, x^F} \int_0^{x^S} O^S(x)dx + \mathbb{E}_\Omega[\sum_{t=1\dots T} \frac{1}{T} \int_0^{x^F_{t,\omega}} O^F(x)dx] \tag{4.1}$$

$$s.t. \quad x_{t,\omega} = x^S + x^F_{t,\omega} \quad \forall t \in \{1 \dots T\}, \forall \omega \in \Omega. \tag{4.2}$$

In this model, the system operator faces an uncertain demand in stage t and scenario ω, $x_{t,\omega}$. The scenario ω belongs to the uncertainty set Ω The TSO has at its disposal an inverse supply curve for fast and slow reserve, $O^F(\cdot)$ and

[3] The general problem with ramp constraints and direct mFRR activation is a multi-stage stochastic program.

$O^S(\cdot)$.[4] The inverse supply curves represent continua of price-taking fast and slow agents.

This problem characterizes activation levels for slow reserve, x^S, and for fast reserve in subperiod t and scenario ω, $x^F_{t,\omega}$, for every combination of inverse supply functions belonging to the set of inverse supply functions, $\mathcal{O}$. This set $\mathcal{O}$ is composed of the positive strictly monotonic increasing functions between arbitrary capacity level, $-N^F$ and N^F, and $-N^S$ and N^S. More specifically, the least-cost activation problem can be interpreted as a social choice function $f : \mathcal{O}^2 \to A$ that maps a combination of inverse supply curves to a dispatching schedule for slow and fast reserve.[5]

The formulation can be made more compact by assuming the subperiod demands to be independently drawn from a random variable, X. This allows us to aggregate the sum over the subperiods as a single realization of uncertainty and to derive the following first-order condition for the optimal activation level for slow reserve:[6]

$$O^S(x^S) = \mathbb{E}_X[O^F(x - x^S)] \tag{4.3}$$

This condition states that the marginal cost of slow reserve should be equal to the expected marginal cost of fast reserve.

4.2.2 Sequence of Markets

The system operator does not have direct control over the flexible assets and needs to organize balancing energy auctions to collect the technical parameters of the assets composing the inverse supply curves. The system operator conducts an *mFRR balancing energy auction* every fifteen minutes to activate mFRR balancing energy over the next fifteen minutes and *aFRR balancing energy auctions* every four seconds to continuously cover the leftover imbalance with aFRR balancing energy.

Alternatively to these centralized auctions, assets can participate directly in the balancing process through imbalance settlement for self-dispatching resources. Agents can self-dispatch to put themselves intentionally in imbalance relative to their market position and thus receive the imbalance price.

This sequence of markets is displayed in figure 4.1 and generates three types of price: aFRR balancing energy prices, P^{aFRR}, the mFRR balancing energy price, P^{mFRR}, and the imbalance price, P^{imb}. The three stages are explained hereunder.

1. Every agent forming the inverse supply curves for fast and slow capacity submits price-quantity balancing energy bids for the aFRR or mFRR

[4] The inverse supply functions are often referred to as merit orders in the power systems literature.

[5] See chapter 23 of [Mas-Colell et al., 1995] for a definition of social choice functions.

[6] The independence assumption might not hold in practice as pointed out in [Papavasiliou et al., 2018]. The results of this paper are not dependent on this assumption but it simplifies the notation.

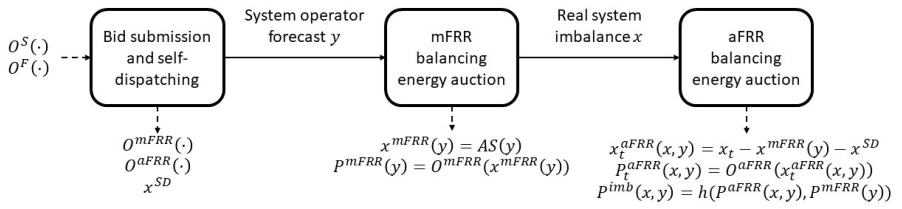


Figure 4.1: Sequence of markets and revelation of uncertainty.

balancing energy auctions or decides to self-dispatch. The aggregations of the balancing energy bids forms the aFRR and mFRR merit order, $O^{aFRR}(\cdot)$ and $O^{mFRR}(\cdot)$, and the level of self-dispatching, x^{SD}.

2. The system operator forecast of the system imbalance, y, is then revealed. The system operator activates a level of mFRR balancing energy, x^{mFRR}, based on this forecast and its mFRR activation strategy, $AS(\cdot)$. The mFRR price follows from the mFRR balancing energy cleared and the mFRR merit order.

3. The system imbalance that the system is actually confronted with, x, is then revealed and the leftover imbalance is covered by aFRR balancing energy, x^{aFRR}. Notice that the balancing constraints include a self-dispatching component that is absent in (4.2). In practice, the system operator can only measure leftover imbalances and the initial system imbalance is indistinguishable from the self-dispatching of agents. The aFRR balancing energy prices follow from the cleared aFRR balancing energy and the aFRR merit order. The imbalance price is then formed by combining the aFRR and mFRR balancing energy price through the imbalance pricing scheme, $h(\cdot)$.

At the second stage, the system operator could follow the least-cost activation strategy defined in (4.1) and (4.2) but it could also activate less or more mFRR than the least-cost activation strategy. TenneT, the Dutch TSO, underlines that it is a "reactive" TSO that "only activates balancing products if imbalances actually occur, not in response to forecasted imbalances."[7] This approach naturally leads to less proactive activation of slow reserve. It is driven by the desire to avoid influencing the system by arbitrarily activating mFRR balancing energy. This approach also aims at preventing dispatching "mistakes" that lead to a counter-activation of aFRR resources to deal with overshooting mFRR activations. Similarly, German TSOs do not consider the economic trade-off between the use of aFRR and mFRR in their mFRR activation strategy [CONSENTEC, 2022]. Activation strategies that lead to an over-activation

[7] See TenneT's website: `https://www.tennet.eu/markets/market-news/balancing-markets`.

of mFRR balancing energy are less frequent so this manuscript focuses on the least-cost mFRR activation strategy and on strategies that activate less mFRR than the least-cost amount.

Both the imbalance pricing scheme and the mFRR activation strategy have an impact on the bidding behaviour of flexible agents. Agents will account for these features as they weigh their options in balancing energy and imbalance payoffs. The activation strategy directly affects the aFRR and mFRR balancing price. The more mFRR that is activated, the higher the mFRR balancing energy price and the lower the aFRR balancing energy price. The effect of the imbalance pricing scheme is more subtle as it can induce self-dispatching. Self-dispatching is reflected in the aFRR and mFRR merit orders as assets that are not available for the balancing energy auctions. It also affects the aFRR balancing energy prices by reducing the leftover imbalances.

The gradual revelation of uncertainty is represented as a realization of a random variable that affects the conditional distribution of imbalances. At bid submission, the initial distribution of imbalance, X, is quite wide but it becomes narrower as the system operator forecast, y, is revealed. The system operator forecast is drawn from a random variable Y and the realization of imbalance that is conditional on y is drawn from the random variable $X|y$.

4.2.3 Individual Bidding Strategy

The strategy set of a fast agent with marginal cost θ is characterized as $\mathcal{Q}^F(\theta)$, the concatenation of the price-quantity aFRR and mFRR balancing energy bids, (p^{aFRR}, q^{aFRR}) and (p^{mFRR}, q^{mFRR}) and the level of self-dispatching, q^{SD}.

$$\mathcal{Q}^F(\theta) = \{(p^{aFRR}, q^{aFRR}, p^{mFRR}, q^{mFRR}, q^{SD}) | q^{aFRR} + q^{mFRR} + q^{SD} = 1, \\ q^{aFRR} \geq 0, q^{mFRR} \geq 0, q^{SD} \geq 0\} \tag{4.4}$$

The objective of the agent is to maximize its payoff from participating in the aFRR balancing energy auction, the mFRR balancing energy auction, and self-dispatching. This profit maximization problem is presented in (4.5) and (4.6):

$$\max \quad q^{aFRR} \cdot z^{aFRR}(\theta, p^{aFRR}) + q^{mFRR} \cdot z^{mFRR}(\theta, p^{mFRR}) + q^{SD} \cdot z^{SD}(\theta) \tag{4.5}$$

$$s.t. \quad (p^{aFRR}, q^{aFRR}, p^{mFRR}, q^{mFRR}, q^{SD}) \in \mathcal{Q}^F(\theta) \tag{4.6}$$

Here, $z^{aFRR}(\theta, p^{aFRR})$ and $z^{mFRR}(\theta, p^{mFRR})$ indicate the payoffs from participating in the aFRR and mFRR balancing auctions respectively, as a function of their price bids, while $z^{SD}(\theta)$ indicates the payoff from self-dispatching.

As the agents are infinitesimal, the payoffs for participating in the balancing energy auctions are not affected by their capacity allocation and are only functions of the price bids that are submitted. The mFRR activation payoff

can be computed through the profit maximization problem (4.7) as a function of the cumulative distribution function of the mFRR balancing energy price, $F_{P^{mFRR}}$:

$$z^{mFRR}(\theta, p^{mFRR}) = \max_{p^{mFRR}} \int_{x \geq p^{mFRR}} (x - \theta) dF_{P^{mFRR}}(x) \tag{4.7}$$

The objective function results from a uniform price auction where an agent is selected as soon as the price of the auction exceeds the price component of the balancing energy bid of the agent. The payoff of the agent is equal to the expectation of the mFRR price minus the marginal cost of the agent whenever the offer is selected. Accepted bids are assumed to be fully selected. Bidding at marginal cost is a weakly dominant strategy in this context and agents effectively act as price-takers. The marginal mFRR activation payoff can then be rewritten as the expectation over the maximum operator between the profit when being activated and zero:

$$z^{mFRR}(\theta) = \int_{x \geq \theta} (x - \theta) dF_{P^{mFRR}}(x) \tag{4.8}$$

$$= \mathbb{E}[\max(P^{mFRR} - \theta; 0)]. \tag{4.9}$$

A similar analysis can be performed to show the weak dominance of bidding at marginal cost in the aFRR balancing auction. This allows us to restrict the strategy set to the capacity components of the bids and the self-dispatching variable.

$$\mathcal{Q}^F(\theta) = \{(q^{aFRR}, q^{mFRR}, q^{SD}) | q^{aFRR} + q^{mFRR} + q^{SD} = 1, \\ q^{aFRR} \geq 0, q^{mFRR} \geq 0, q^{SD} \geq 0\} \tag{4.10}$$

The marginal self-dispatching payoff is not based on an energy auction but rather on the self-activation of an agent based on its expectation of the imbalance price. The payoff can be found by maximizing the expected payoff of performing self-dispatching, ai, given the expected imbalance price. This is expressed in problem (4.11):

$$z^{SD}(\theta) = \max_{ai} (\mathbb{E}[P^{imb}] - \theta) \cdot ai \tag{4.11}$$

$$s.t. \quad 0 \leq ai \leq 1$$

We assume that the decision on how much capacity to allocate to self-dispatching is made at gate closure of the mFRR/aFRR balancing energy auctions, before y is revealed. We directly see that $ai = 0$ if $\mathbb{E}[P^{imb}] \leq \theta$ and $ai = 1$ otherwise. This means that an agent should commit to self-dispatching its asset only if the expected imbalance price is greater than its marginal cost. This allows us to reformulate the self-dispatching payoff as

$$z^{SD}(\theta) = \max(\mathbb{E}[P^{imb}] - \theta; 0). \tag{4.12}$$

The allocation of the capacity between aFRR and mFRR balancing energy auctions and self-dispatching then depends on the ranking of the payoff of the different options which are dependent on the aFRR and mFRR balancing energy price and the imbalance price. The strategy set for a slow agent with marginal cost θ is defined similarly as $\mathcal{Q}^S(\theta)$ except that there is no aFRR component.

$$\mathcal{Q}^S(\theta) = \{(q^{mFRR}, q^{SD}) | q^{mFRR} + q^{SD} = 1, q^{mFRR} \geq 0, q^{SD} \geq 0\} \quad (4.13)$$

4.2.4 Aggregated Bidding Strategy

Aggregating the individual allocation strategies generates *aggregate functions* belonging to aggregate strategy sets. To be more specific, an aggregate strategy belonging to the aggregate strategy set for fast agents, $\mathcal{Q}^F$, is composed of the aggregate function for the aFRR balancing energy auction, $q^{aFRR}(\cdot)$, the aggregate function for the mFRR balancing energy auction, $q_F^{mFRR}(\cdot)$, and the aggregate function for self-dispatching, $q_F^{SD}(\cdot)$. These aggregate functions describe the aggregate strategy of fast agents as a function of their marginal cost. Similarly, an aggregate strategy for slow agents belongs to the aggregate strategy set for slow agents, $\mathcal{Q}^S$, and is composed of the aggregate function for participating in the mFRR balancing energy auction, $q_S^{mFRR}(\cdot)$, and the aggregate function for self-dispatching, $q_S^{SD}(\cdot)$.

$$(q^{aFRR}(\cdot), q_F^{mFRR}(\cdot), q_F^{SD}(\cdot)) \in \mathcal{Q}^F \quad (4.14)$$

$$(q_S^{mFRR}(\cdot), q_S^{SD}(\cdot)) \in \mathcal{Q}^S \quad (4.15)$$

The conversion between aggregate strategy sets and merit orders to match the notation of figure 4.1 is provided analytically in (4.16) to (4.18) for aFRR, mFRR and self-dispatching allocation functions given cumulative distribution functions for fast and slow capacity, G_F and G_S, and for capacity levels, N^F and N^S:

$$(O^{aFRR})^{-1}(\theta) = N^F \int_0^\theta q^{aFRR}(x) dG_F(x) \quad (4.16)$$

$$(O^{mFRR})^{-1}(\theta) = N^F \int_0^\theta q_F^{mFRR}(x) dG_F(x) + N^S \int_0^\theta q_S^{mFRR}(x) dG_S(x) \quad (4.17)$$

$$x^{SD} = N^F \left(\int q_F^{SD}(x) dG_F(x) + N^S \int q_S^{SD}(x) dG_S(x) \right) \quad (4.18)$$

The supply curve for aFRR is defined in (4.16) by integrating the capacity of the fast agents with a marginal cost lower than θ that allocate their capacity to the aFRR balancing energy auction. The supply curve for mFRR is defined similarly except that both fast and slow agents can submit mFRR

balancing energy bids. The level of self-dispatching is obtained by integrating slow and fast capacity that allocates its capacity to self-dispatching. Note that the supply curves for fast and slow capacity can be also be derived analytically:

$$(O^S)^{-1}(\theta) = Q \int_0^{\theta} 1 dG_S(x) = Q \cdot G_S(\theta) \tag{4.19}$$

$$(O^F)^{-1}(\theta) = Q \int_0^{\theta} 1 dG_F(x) = Q \cdot G_F(\theta) \tag{4.20}$$

4.3 Equilibrium

This section describes the equilibria that are generated by different combinations of mFRR activation strategies, AS, and imbalance pricing schemes, h. Each combination is modeled as a mechanism

$$\Gamma(AS, h) = \{\mathcal{Q}^F, \mathcal{Q}^S, g(\cdot|AS)\} \tag{4.21}$$

where $g(\cdot|AS) : \mathcal{Q}^F \times \mathcal{Q}^S \to A$ is an outcome function that maps aggregated strategy sets for fast and slow capacity to a dispatching schedule given an mFRR activation strategy. Two types of activation strategies are investigated: (i) the least-cost activation strategy that replicates the mFRR dispatch of the social optimum in the two-stage stochastic program, and (ii) every mFRR activation strategy activating less mFRR than the least-cost ideal. Three imbalance pricing schemes are investigated: (i) the "mFRR only" imbalance price inspired by the "law of one price", (ii) the "mean mFRR and aFRR" and (iii) the "max mFRR and mean aFRR" imbalance price as stated by article 9 of the ISHM. This results in six possible cases that can be restricted to four equilibria: (1) the "mFRR only" imbalance price for both mFRR activation strategies, (2) the "mean mFRR and aFRR" imbalance price with the least-cost mFRR activation strategy, (3) the "mean mFRR and aFRR" imbalance price with an activation of mFRR which is less than that of the least-cost mFRR activation strategy, and (4) the "max mFRR and mean aFRR" imbalance price for both activation strategies. The equilibria are summarised in table 4.1 and consist of (1) every slow agent participating in the mFRR balancing energy auction and every fast agent participating in the aFRR balancing energy auction for the "mFRR only" price, (2) same for the "mean mFRR and aFRR" imbalance pricing method with the least-cost activation strategy, (3) some slow agents self-dispatching and others participating in the mFRR balancing energy auction and every fast agent participating in the aFRR balancing energy auction for the "mean mFRR and aFRR" imbalance pricing method with less mFRR activation than that of the least-cost mFRR activation strategy, and (4) some fast and slow agents self-dispatching and others participating either in the aFRR or mFRR balancing energy auction for the "max mFRR and mean mFRR" imbalance pricing scheme.

The remainder of the section will cover the different cases.

Table 4.1: Summary of equilibria for different combinations of imbalance pricing schemes and activation strategies.

Imb. Pricing Schemes	mFRR only		mean mFRR and aFRR		max mFRR and mean aFRR	
Asset / Act. Strat.	Slow	Fast	Slow	Fast	Slow	Fast
Least-cost	mFRR	aFRR	mFRR	aFRR	self-disp. and mFRR	self-disp. and aFRR
Less mFRR			self-disp and mFRR	aFRR		

4.3.1 "mFRR only"

The "mFRR only" imbalance pricing scheme is motivated by the "law of one price": homogeneous goods should trade at the same price [Jevons, 1871]. This imbalance pricing scheme considers that imbalances and mFRR balancing energy are at least partially substitutable, based on the fact that they both represent balancing energy that is traded on a 15-minute timescale and should thus be priced similarly. This results in the imbalance price aligning to the mFRR price:

$$h(P^{aFRR}(x,y), P^{mFRR}(y)) = P^{mFRR}(y). \tag{4.22}$$

One property of this scheme is that no agents find it to their advantage to self-dispatch, since participating in the mFRR balancing auction is always more profitable.

Lemma 4.1. *The payoff from participating in the mFRR balancing auction is greater than or equal to the one from self-dispatching under the "'mFRR only" imbalance price.*

Proof. By Jensen's inequality on (4.9) and (4.12),

$$\max(\mathbb{E}[P^{imb}]-\theta;0) = \max(\mathbb{E}[P^{mFRR}]-\theta;0) \leq \mathbb{E}[\max(P^{mFRR}-\theta;0)]. \tag{4.23}$$

□

We can now analyze the impact of the mFRR activation strategy on the payoff of agents. We begin by applying the least-cost activation strategy to the aFRR and mFRR merit order curves.

Lemma 4.2. *The payoff from participating in the mFRR balancing energy auction is lower than or equal to the payoff of participating in the aFRR balancing energy auction if the mFRR activation strategy of the system operator follows the least-cost strategy.*

Proof. By Jensen's inequality,

$$\max(P^{mFRR} - \theta; 0) = \max(\mathbb{E}[P^{aFRR}] - \theta; 0) \leq \mathbb{E}[\max(P^{aFRR} - \theta; 0)]. \quad (4.24)$$

□

The difference between the aFRR and the mFRR payoffs can be considered as a flexibility premium. We are now ready to characterize the equilibrium under the "mFRR only" imbalance price.

Proposition 4.3. *Every agent offering their capacity to the best quality auction they can (fast agents to aFRR balancing energy and slow agents to mFRR balancing energy), which corresponds to the aggregate functions*

$$(q^{aFRR*}(\cdot), q_F^{mFRR*}(\cdot), q_F^{SD*}(\cdot)) = (1, 0, 0) \quad \textit{and} \quad (4.25)$$

$$(q_S^{mFRR*}(\cdot), q_S^{SD*}(\cdot)) = (1, 0), \quad (4.26)$$

is a dominant strategy equilibrium under the "mFRR only" imbalance price and with an mFRR activation strategy equal to the least-cost activation strategy. This mechanism implements the least-cost social choice function in dominant strategy.[8]

Proof. Lemmas 4.1 and 4.2 state that, for all agents, performing self-dispatching is less profitable than participating in the mFRR balancing energy auction and that the aFRR balancing energy auction is more profitable than the mFRR balancing energy auction. Every slow agent offering their capacity to the mFRR balancing energy auction and every fast agent offering their capacity to the aFRR balancing energy auction are dominant strategies. □

This results in aFRR prices following the inverse supply curve of fast agents and mFRR prices following the inverse supply curve of slow agents. The analysis for mFRR activation strategies with less mFRR activation than the least-cost mFRR activation is trivial, since lemma 4.1 is not affected by the mFRR activation strategy and lemma 4.2 remains valid for mFRR activation strategies that activate less mFRR capacity than that activated in the least-cost mFRR activation strategy. The only difference in that context would be that the mFRR activation strategy prevents the balancing markets from reaching the least-cost allocation.

4.3.2 "mean mFRR and aFRR" Imbalance Settlement with Least-Cost Activation Strategy

The "mean mFRR and aFRR" imbalance price is expressed in equation (4.27) as a function of the demands for mFRR and aFRR balancing energy, x^{mFRR} and x_t^{aFRR} for $t \in \{1 \dots T\}$:

$$h(P^{aFRR}(x, y), P^{mFRR}(y)) = \frac{\sum_{t=1\dots T} \frac{1}{T} P_t^{aFRR}(x_t, y) + P^{mFRR}(y)}{2} \quad (4.27)$$

[8] See chapter 23 of [Mas-Colell et al., 1995] for an introduction to mechanism design.

Proposition 4.4. *Every agent offering their capacity to the best quality auction they can is a dominant strategy equilibrium under the "mean mFRR and aFRR" imbalance price and with an mFRR activation strategy equal to the least-cost activation strategy. This mechanism implements the least-cost social choice function in dominant strategy.*

Proof. Under the least-cost mFRR activation strategy, the mFRR price is equal to the expected aFRR price, $P^{mFRR} = \mathbb{E}[P^{aFRR}]$, which establishes an expected imbalance price equal to the expected mFRR price (see equation (4.27)). Lemmas 4.1 and 4.2 can then both be applied to prove the desired result. □

Interestingly, both propositions 4.3 and 4.4 can also be interpreted as competitive equilibria. No agent has any incentive to deviate from their aggregate function given the aFRR, mFRR, and imbalance prices that are generated.

4.3.3 "mean mFRR and aFRR" Imbalance Settlement with Less mFRR than Least Cost

If the mFRR activation strategy activates less than the amount of mFRR activated by the least-cost activation strategy, self-dispatching may become more profitable than participating in the mFRR balancing energy auction. This is due to the higher aFRR prices that lift the imbalance price. This results in the following proposition.

Proposition 4.5. *Every agent offering their capacity to the best quality auction they can is not always an equilibrium under the "mean mFRR and aFRR" imbalance price and with an mFRR activation strategy that activates less mFRR than the least-cost strategy.*

An example of truthful participation in the balancing energy auction not being an equilibrium, proving proposition 4.5, is provided in section 4.4.2. The equilibrium in that case results in self-dispatching from the slow agents and is characterized by the following aggregate function:

- Every fast agent participates in the aFRR balancing energy auction:

$$(q^{aFRR*}(\cdot), q_F^{mFRR*}(\cdot), q_F^{SD*}(\cdot))) = (1, 0, 0) \tag{4.28}$$

- Slow agents with marginal cost lower than $O^S(x_S^{SD})$, i.e. the cheapest capacity up to a quantity x_S^{SD}, self-dispatch and other slow agents decide to participate in the mFRR balancing energy auction:

$$(q_S^{mFRR*}(\theta), q_S^{SD*}(\theta)) = \begin{cases} (0,1) & \text{if } O^S(0) \leq \theta \leq O^S(x_S^{SD}), \\ (1,0) & \text{else.} \end{cases} \tag{4.29}$$

A level x_S^{SD} of self-dispatch from slow agents has two direct effects: it increases the mFRR prices by removing cheap assets from the mFRR merit order,

and it decreases the aFRR prices by reducing the leftover imbalance that needs to be covered by aFRR. The specific structure of these aggregate functions allows us to derive a closed-form formulation for the mFRR merit orders without solving (4.17) explicitly. Generators with a marginal cost lower than $O^S(x_S^{SD})$ opt out of the mFRR balancing energy auction and this translates the inverse supply function of slow assets to the left by x_S^{SD} for upward balancing capacity:

$$O^{mFRR}(x) = \begin{cases} O^S(x + x_S^{SD}) & \text{if } x \geq 0, \\ O^S(x) & \text{else.} \end{cases} \tag{4.30}$$

As shown in figure 4.1, self-dispatching decreases the demand for aFRR balancing energy and the subsequent aFRR prices. The demand for mFRR balancing is not affected by the self-dispatching, as this demand is obtained from the system operator forecast and an mFRR activation strategy that is assumed to be unaffected by the self-dispatching. Increasing the self-dispatch level increases the mFRR activation payoff, due to the increased mFRR prices, and decreases the self-dispatching payoff, due to the decreased aFRR prices. An equilibrium can then be found by finding a level of self-dispatch x_S^{SD} such that every agent below x_S^{SD} on the inverse supply curve finds self-dispatching more profitable, every agent after x_S^{SD} finds participating in the mFRR balancing energy auction more profitable, whereas the agent at x_S^{SD} is indifferent between both options. This corresponds to solving the following identity where the payoff for both mFRR activation and self-dispatch are dependent on the level of self-dispatching:

$$z^{mFRR}(O^S(x_S^{SD})|x_S^{SD}) = z^{RB}(O^S(x_S^{SD})|x_S^{SD}). \tag{4.31}$$

4.3.4 "max mFRR and mean aFRR" Imbalance Pricing

The "max mFRR and mean aFRR" imbalance price as a function of the demands for mFRR and aFRR balancing energy is characterized as follows:

$$h(P^{aFRR}(x,y), P^{mFRR}(y)) = \max\left(\sum_{t=1\ldots T} \frac{1}{T} P_t^{aFRR}(x_t, y); P^{mFRR}(y)\right) \tag{4.32}$$

Under this imbalance settlement scheme, the expected imbalance price is always greater than the expected aFRR price or the mFRR price. Both fast and slow agents may find it optimal to self-dispatch at equilibirum. As in the case of the "mean mFRR and aFRR" price, self-dispatch is performed by the cheapest agents.

Proposition 4.6. *Every agent offering their capacity to the best quality auction they can is not always an equilibrium under the "max mFRR and mean aFRR" imbalance price.*

The difference between proposition 4.5 and 4.6 is that even the least-cost activation strategy does not always sustain an equilibrium where every agent offers their capacity to the best quality auction. An example proving 4.6 is also given in section 4.4.2. The equilibrium in this case exhibits self-dispatching from both fast and slow agents, and can be characterized as follows:

- Fast agents with marginal cost lower than $O^F(x_F^{SD})$ self-dispatch and other fast agents decide to participate in the aFRR balancing energy auction.

$$(q^{aFRR*}(\theta), q_F^{mFRR*}(\theta), q_F^{SD*}(\theta)) = \begin{cases} (0,0,1) & \text{if } O^F(0) \leq \theta \leq O^F(x_F^{SD}), \\ (1,0,0) & \text{else.} \end{cases} \tag{4.33}$$

- As in (4.29), slow agents with marginal cost lower than $O^S(x_S^{SD})$ self-dispatch and others decide to participate in the mFRR balancing energy auction.

Self-dispatching by fast assets is not an equilibrium outcome in the "mean mFRR and aFRR" case as the expected imbalance price is bounded by the mean mFRR and aFRR prices and if the mean mFRR price becomes greater than the mean aFRR price, self-dispatching becomes strictly less profitable than offering mFRR. For the "max mFRR and mean aFRR" imbalance pricing design, if the mFRR price becomes greater than the aFRR price and dominates the formation of the imbalance price, self-dispatching is as profitable as offering mFRR and can be strictly greater than offering aFRR. This can result in self-dispatching from both fast and slow assets. The equilibrium can be obtained by finding the fast and slow agents that are indifferent between self-dispatching and offering aFRR or mFRR, depending on the level of fast and slow self-dispatching, x_F^{SD} and x_S^{SD}. These thresholds are characterized by the following identity:

$$z^{mFRR}(O^S(x_S^{SD})|x_S^{SD}) = z^{RB}(O^S(x_S^{SD})|x_S^{SD}, x_F^{SD}), \tag{4.34}$$

$$z^{aFRR}(O^F(x_F^{SD})|x_S^{SD}, x_F^{SD}) = z^{RB}(O^F(x_F^{SD})|x_S^{SD}, x_F^{SD}). \tag{4.35}$$

4.4 Results

The results that are presented in this section are based on an illustrative example. Let $N^F = 500$ MW and $N^S = 1000$ MW be the fast and slow capacity that is available for balancing. The aggregate marginal cost of fast and slow balancing capacity is distributed uniformly between 0 and 100 €/MWh. Let $Y(w)$ be the forecast of the system operator be uniformly distributed between $-100 + w$ and $100 + w$. w can be interpreted as the forecast at the *gate-closure* of the aFRR and mFRR balancing energy auction. It is known before the system operator forecast, y, is revealed and before the formation of the

equilibrium. For each realization of the gate-closure forecast, w, there is a corresponding equilibrium.Let $X|y$ be the random variable from which the actual system imbalance is drawn. It is uniformly distributed between $y-100$ and $y+100$.

This section begins by describing the least-cost mFRR activation strategy in this setting. It continues with the equilibria for the different imbalance pricing schemes, illustrating propositions 4.5 and 4.6. Finally, we focus on the activation cost under the different pricing schemes and mFRR activation strategies.

4.4.1 Least-Cost mFRR Activation Strategy

The least-cost mFRR activation strategy is obtained by expressing the activation of slow capacity as a function of the system operator forecast of the imbalance, y, as indicated in condition (4.3). It results in the following identity when applied to our numerical settings:

$$O^S(x^S) = \int O^F(x-x^S)dF_{X|y}(x) = \int_{y-100}^{y+100} O^F(x-x^S)\frac{1}{200}dx. \tag{4.36}$$

Given $O^S(x) = x/10$ and $O^F(x) = x/5$, the least-cost activation strategy, $AS^*(y)$, can be expressed analytically as:

$$x^S = \frac{2}{3}y = AS^*(y). \tag{4.37}$$

This strategy is also used to determine the demand for mFRR balancing energy by replacing the fast and slow merit orders by the aFRR and mFRR merit orders.

4.4.2 Self-Dispatching

Figure 4.2 presents the capacity that is not offered to balancing energy auctions as a function of w for the illustrative example. A new equilibrium with its associated level of self-dispatch is established for every realization of w. This figure presents the level of self-dispatching for a mFRR activation strategy that activates less mFRR than the least-cost activation strategy,

$$AS(y) = 0.5 \cdot y < AS^*(y), \tag{4.38}$$

and for the imbalance pricing schemes that generate self-dispatch, namely the "mean mFRR and aFRR" and the "max mFRR and mean aFRR" imbalance pricing schemes. The illustrative example is symmetric and exhibits a similar level of self-dispatching by downward flexible assets for negative w. Higher gate-closure forecasts indicate a higher discrepancy between the aFRR and mFRR balancing energy prices and result in higher self-dispatching payoffs for imbalance pricing schemes that include an aFRR price component. There is a

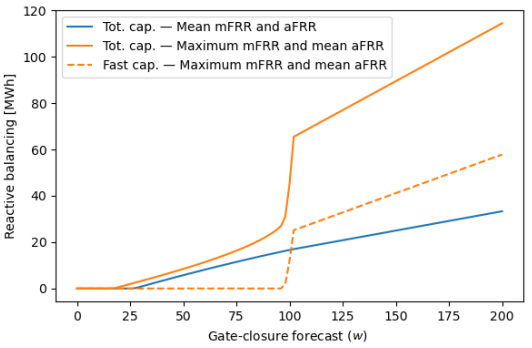


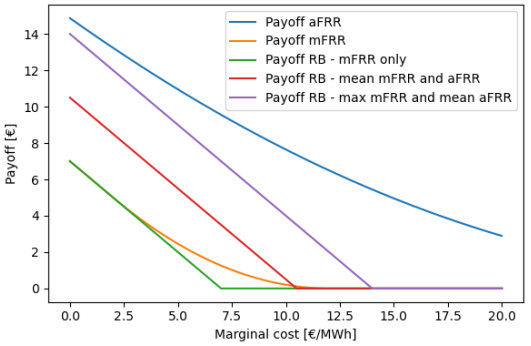
Figure 4.2: Fast and slow assets self-dispatching when the mFRR activation strategy is lower than that of the least-cost activation strategy ($AS(y) = 0.5 \cdot y$).

change of mode for the "max mFRR and aFRR" imbalance price when the gate-closure forecast reaches around 100 MW. Only slow agents self-dispatch before 100 MW, but, as soon as the forecast reaches 100 MW, fast agents start to self-dispatch and this increases the total capacity not offered to the balancing energy auctions. This coincides with the mFRR price dominating the mean aFRR price in the formation of the imbalance price.

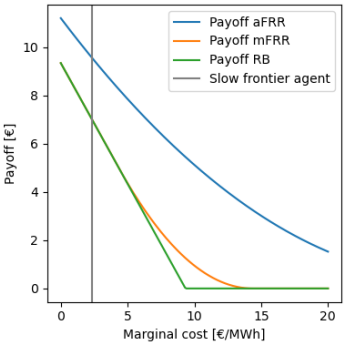
Figure 4.3 illustrates the effect of self-dispatching on the expected payoffs of the agents as a function of their marginal cost for a given gate-closure forecast w. Figure 4.3a shows that, under the mFRR activation strategy $AS(y) = 0.5 \cdot y$ and the gate-closure forecast $w = 140$, every fast asset offering aFRR and every slow asset offering mFRR cannot constitute an equilibrium for either the "mean mFRR and aFRR" or the "max mFRR and mean aFRR" imbalance prices. The mFRR balancing energy payoff is dominated by the self-dispatching payoff, and agents would rather self-dispatch than participate in the mFRR balancing energy auctions.

Figure 4.3b presents the payoffs under the equilibrium level of self-dispathcing for the "mean mFRR and aFRR" imbalance price. Slow assets that self-dispatch reduce the aFRR prices and payoffs (by reducing the residual system imbalance that needs to be covered by aFRR) and increase the mFRR prices and payoffs (by removing cheap assets from the mFRR merit order). Slow assets self-dispatch up to the point where the expected incremental payoff of self-dispatching and of participating in the mFRR balancing energy auction become equal. This characterizes a frontier agent such that all slow assets with lower marginal cost self-dispatch and all assets with higher marginal cost participate in the mFRR balancing energy auction.

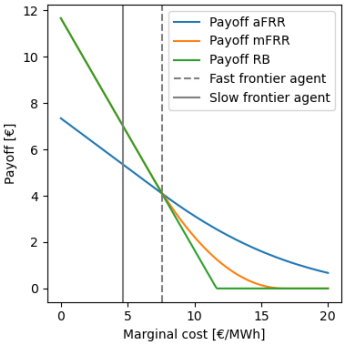
Figure 4.3c provides a similar analysis for the "max mFRR and mean aFRR" imbalance price. The self-dispatching from slow assets results in an increased mFRR price dominating the mean aFRR price in the formation of the imbalance price. This leads to a large level of self-dispatching from fast assets.



(a) No self-dispatch.



(b) Optimal level of self-dispatching for "mean mFRR and aFRR".



(c) Òptimal level of self-dispatching for "max mFRR and mean aFRR".

Figure 4.3: Activation payoffs for $w = 140$ and $AS(y) = 0.5 \cdot y$.

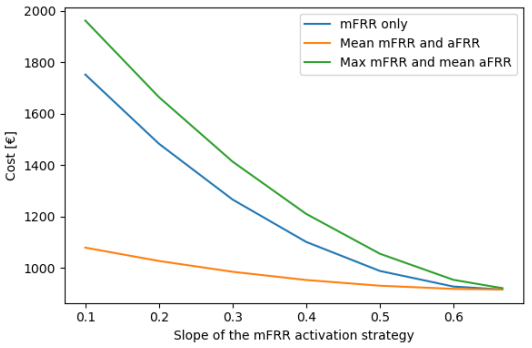


Figure 4.4: Activation cost in the illustrative example as a function of the mFRR activation strategy.

4.4.3 Activation cost

Figure 4.4 presents the activation that results from different mFRR activation strategies and imbalance pricing schemes. The minimum activation cost can be found under the least-cost mFRR activation strategy for the "mFRR only" and the "mean mFRR and aFRR" imbalance settlement schemes. The "mFRR only" imbalance price generates the benchmark activation cost where all agents offer their capacity to the best quality auction they can. The "mean mFRR and aFRR" can reduce the activation cost by correcting an inefficient mFRR activation strategy. It incentivizes slow agents to self-activate and it drives the equilibrium closer to the optimal least-cost dispatch. The self-activation resulting from the "max mFRR and aFRR" imbalance pricing scheme is inefficient and over-compensates for the inaccurate mFRR activation strategy.

4.5 Discussion

This section discusses the policy implications of the results that are illustrated in the previous section and the impact of some modeling assumptions.

4.5.1 Bidding Incentives

One finding of our analysis is that bidding the marginal cost in the balancing energy auctions is not always the optimal strategy for the "mean mFRR and aFRR" and the "max mFRR and mean aFRR" imbalance settlement schemes. The "mFRR only" imbalance settlement scheme has the advantage of providing a clear *weakly dominant* strategy for the agents: they should bid in the best-quality auction they can. An advantage of a clear optimal bidding strategy is that it fosters competition by reducing the barrier to entry. There is no need to rely on extended analytics in order to participate profitably in the market.

The other imbalance pricing schemes can allow for pure strategy equilibria, but they can be harder to reach in practice. They require agents to perfectly anticipate the behaviours of the other agents, since their strategy depends on the level of self-dispatching in the system.

The analysis also highlights the influence of arbitrary components, such as the system operator forecast, y, on the bidding incentives. Reacting purely to the imbalance and not proactively activating mFRR has an impact on the market as it influences the mFRR and aFRR prices and payoffs. It also propagates to the imbalance prices and payoffs.

4.5.2 Balancing Process Efficiency

The objective of this analysis is not to condemn self-dispatching as a practice but rather to isolate and highlight the balancing energy market features that support an efficient balancing process. Balancing market features that provide incentives for self-dispatching can have a positive social impact if TSOs do not activate optimally as seen by the reduction in balancing cost induced by the "mean mFRR and aFRR" imbalance prices.[9] However, these features should not unduly influence the balancing energy auctions or impede the efficiency of the balancing process. The analysis in the paper has shown that some imbalance pricing schemes (here the "max mFRR and mean aFRR") cannot support an efficient dispatch.

The connection to the cross-border balancing platforms has sparked a debate on self-dispatching incentives. ELIA, the Belgian TSO, wants to avoid pricing signals that may aggravate its own position to relieve the aggregated European imbalance [ELIA, 2023]. CREG, the Belgian regulator, believes that prices should support self-dispatching that helps the aggregated European imbalance [CREG, 2023]. This debate can be avoided with the "mFRR only" imbalance pricing scheme that provides incentives for flexible agents to offer their capacity to the balancing energy markets. It can help to fully harness the coordination benefits brought forth by the connection to the platforms.

4.5.3 Robustness to Modelling Assumptions

In practice, balancing markets are split into upward and downward markets. The simultaneous activation of balancing energy in both directions can result from ramp constraints. For example, if some upward balancing energy is activated but the imbalance decreases and the upward balancing energy cannot decrease fast enough, downward balancing energy may be activated to keep the balance. The simultaneous activation of both products is dealt with in practice through separate markets. In the absence of ramp constraints, as this paper assumes, it is sufficient to consider a single merit order that extends for both negative and positive balancing energy.

[9] Self-dispatching can also enable TSOs to access flexible capacity that faces difficulties in participating explicitly in the balancing energy market.

In general, a market design should be evaluated based on its ability to elicit both the type of the agents (fast or slow in our case) and their marginal cost. This analysis only focuses on eliciting the type by assuming price-taking fringe agents. Market power considerations are ignored. There is value in understanding market behavior under simplified assumptions. Designs that fail this test are highly questionable for more complex conditions that include market power. Our analysis shows that the case without market power is already highly non-trivial, and is a necessary first step before venturing into more complex analysis. The assumption on market power is anyway aligned with REMIT, the European approach for tackling market power. REMIT is an ex-process process that checks the competitiveness of the bids after market clearing.

The model assumes that assets commit to self-dispatching at gate closure. This is a realistic representation of some European countries where the system imbalance is revealed with a thirty-minute delay.[10] In other countries, such as Belgium and the Netherlands, the system imbalance is revealed in real time, at every minute. Our model underestimates the benefits of self-dispatching in this context, since the risk associated with self-dispatching decreases with the level of information that agents have on the system imbalance. Nevertheless, the insights of the "mFRR only" imbalance pricing scheme are independent of this assumption. Self-dispatching is always weakly dominated by participating in the mFRR balancing energy auction, regardless of the information that is available to the agent at the time of balancing. If assets can commit to self-dispatching at the beginning of the ISP, or adapt dynamically to the revelation of uncertainty as the ISP unfolds, a qualitative argument can be made for an increased level of self-dispatch for the "mean mFRR and aFRR" and the "max mFRR and mean aFRR" imbalance prices. A quantitative analysis would require the characterization of the equilibrium.

The model also assumes that the system operator demand for mFRR balancing energy is not endogenized. In particular, the activation strategy does not account for the induced self-dispatch. If the system operator is considered to be an agent that participates in the multi-stage game with the objective of minimizing activation cost, then only three cases based on the imbalance pricing scheme need to be inspected. There is no need anymore to consider different activation strategies, since the system operator would submit a demand for mFRR balancing energy that would minimize activation cost. The equilibrium with a responsive system operator and the "mFRR only" and the "mean mFRR and aFRR" prices would actually be identical to the one generated by an unresponsive system operator following the least-cost activation strategy. The case with the "max mFRR and mean aFRR" is not as straightforward, as there may exist an activation strategy such that the induced self-dispatching results in a lower activation cost than the least-cost activation strategy of an unresponsive system operator.

[10] European regulation mandates system operators to publish the system imbalance with a delay of at most thirty minutes.

4.6 Discussion

This section discusses the policy implications of the results that are illustrated in the previous section and the impact of some modelling assumptions.

4.6.1 Policy Implications

One finding of our analysis is that bidding the marginal cost in the balancing energy auctions is not always the optimal strategy for the "mean mFRR and aFRR" and the "max mFRR and mean aFRR" imbalance settlement schemes. The "mFRR only" imbalance settlement scheme has the advantage of providing a clear strategy for the agents: they should bid in the best-quality auction they can, and they have no opportunity cost in the capacity auctions. This strategy is *weakly dominant*, and independent of the other agents' strategies. An advantage of a clear optimal bidding strategy is that it fosters competition by reducing the barrier to entry. There is no need to rely on extended analytics in order to participate profitably in the market. The other imbalance pricing schemes can allow for pure strategy equilibria, but they can be harder to reach in practice. They require agents to perfectly forecast the behaviours of the other agents, since their strategy depends on the level of reactive balancing in the system.

Our analysis also goes against one argument in favor of reactive balancing. It is argued that reactive balancing decreases the capacity procurement cost, by reducing the *reserve requirement*. The reserve requirement drives the width of the capacity demand curve and it is computed based on the historical distribution of the system imbalance, which can be reduced through reactive balancing [ELIA, 2021b]. This reasoning ignores the potential capacity cost increase that is caused by a balancing setting that incentivizes reactive balancing and generates artificial opportunity cost that is driven by imbalance settlement pricing as opposed to the intrinsic economic value of reserve. The more reactive balancing is profitable, the higher the opportunity cost for agents to participate in the balancing energy auctions and the higher the prices in the capacity auctions. It is unclear which factor, the reduction in reserve requirements or the increase in capacity prices, exerts a greater impact on the procurement cost, and this should be evaluated in a system-specific manner, but the argument cited above that has been used in public discourse is incomplete.

4.6.2 Effect of Modelling Assumptions

The model assumes that assets commit to performing reactive balancing at gate closure. This is a realistic representation of some European countries where the system imbalance is revealed with a thirty-minute delay.[11] In other countries, such as Belgium and the Netherlands, the system imbalance is revealed in real

[11] European regulation mandates system operators to publish the system imbalance with a delay of at most thirty minutes.

time, at every minute. Our model underestimates the benefits of performing reactive balancing in this context, since the risk associated with performing reactive balancing decreases with the level of information that agents have on the system imbalance. Nevertheless, the insights of the "mFRR only" imbalance pricing scheme are independent of this assumption. Performing reactive balancing is always weakly dominated by participating in the mFRR balancing energy auction, regardless of the information that is available to the agent at the time of balancing. If assets can commit to performing reactive balancing at the beginning of the ISP, or adapt dynamically to the revelation of uncertainty as the ISP unfolds , a qualitative argument can be made for an increased level of reactive balancing for the "mean mFRR and aFRR" and the "max mFRR and mean aFRR" imbalance prices. A quantitative analysis would require the characterization of the equilibrium.

The model also assumes that the system operator demand for mFRR balancing energy is not endogenized. In particular, the activation strategy does not account for the induced self-dispatch. If the system operator is considered to be an agent that participates in the multi-stage game with the objective of minimizing activation cost, then only three cases based on the imbalance pricing scheme need to be inspected. There is no need anymore to consider different activation strategies, since the system operator would submit a demand for mFRR balancing energy that would minimize activation cost. The equilibrium with a responsive system operator and the "mFRR only" and the "mean mFRR and aFRR" prices would actually be identical to the one generated by an unresponsive system operator following the least-cost activation strategy. The case with the "max mFRR and mean aFRR" is not as straightforward, as there may exist an activation strategy such that the induced self-dispatching results in a lower activation cost than the least-cost activation strategy of an unresponsive system operator.

Finally, the *one-way substitutability* assumption of the fast assets can be discussed. It is common to assume that fast-moving assets can offer both mFRR and aFRR without restriction, however this ignores energy-constrained assets such as batteries or pump-hydro power plants. These assets can have difficulties in participating in the mFRR auction due to the longer activation time in the same direction. Without the one-way substitutability, our model could lead to price reversal in aFRR and mFRR capacity prices.

4.7 Conclusion

This chapter proposes a framework for analyzing European multi-product balancing auctions. We analyze the impact of the imbalance settlement scheme and the mFRR activation strategy on the balancing market equilibria. The reaction of rational fringe agents is endogenously accounted for by the model.

Four main insights can be derived from the model. (1) The "mFRR only" imbalance pricing scheme incentivizes agents to offer their capacity to the best-

quality balancing energy auction they can. (2) The minimum balancing activation cost can be reached with the "mFRR only" and the "'mean mFRR and aFRR" imbalance prices under the least-cost mFRR activation strategy. (3) If the least-cost activation strategy is not applied, the "mean mFRR and aFRR" imbalance price incentivizes a level of reactive balancing compensating for the inefficient activation strategy and generates a lower balancing activation cost than the benchmark equilibrium where agents participate in the best balancing energy auction they can. (4) The "max mFRR and mean aFRR" imbalance price induces a level of reactive balancing that increases the balancing activation cost.

Future works aim to extend the framework to account for cross-border interactions through the European balancing platforms and empirically estimate the mFRR activation strategy of European TSOs. Another line of research will focus on the characterization of elastic demand curves for mFRR balancing energy.

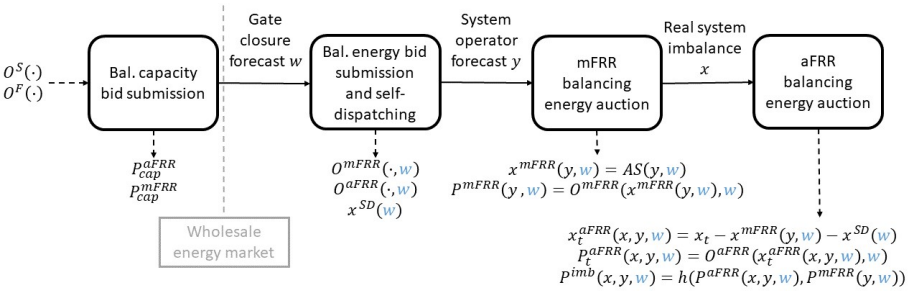


Figure 4.5: Sequence of markets and revelation of uncertainty with balancing capacity markets.

4.A Balancing Capacity Markets

The sequence of markets updated to account for balancing capacity markets is displayed in figure 4.5. The capacity auctions (also called *reserve auctions*) are assumed to be held consecutively.[12] They generate balancing capacity prices, P_{cap}^{aFRR} and P_{cap}^{mFRR}, based on the aFRR and mFRR capacity demand curves and the fast and slow assets balancing capacity bids.

The fast assets that are not selected can compete in the mFRR capacity auction with the slow assets. The capacity that is cleared in either the aFRR or the mFRR capacity auction has to be offered to the corresponding energy auction. The capacity that is not cleared can either participate in one of the energy auctions through so-called free bids or self-dispatch.

Modeling balancing capacity markets requires the introduction of the gate closure forecast, w, drawn from the random variable W. It represents the information available at the gate-closure of the balancing energy auctions and allows for multiple equilibria.[13] Each equilibrium as a function of w can be analyzed independently The gate-closure forecast was already introduced in the result section.

Balancing capacity markets interact with both the wholesale energy market and the balancing energy markets. The need for balancing capacity markets stems from an opportunity cost for participating in the balancing energy markets. If wholesale energy markets are more profitable than balancing energy markets, flexible assets need additional incentives to participate in the balancing energy auctions. This analysis does not explicitly account for the wholesale market but it shows the potential opportunity cost that can be generated by combinations of mFRR activation strategies, imbalance pricing schemes, and capacity demand curves.

[12] The product of higher quality, aFRR, is assumed to be auctioned first in our analysis. The joint co-optimized procurement of aFRR and mFRR balancing capacity is out of scope, although this is a very active field of debate at present, for instance in the UK.

[13] Article 8.2 of the *Implementation Framework for the European Platform for the Exchange of Balancing Energy from Frequency Restoration Reserve with Automatic Activation.*

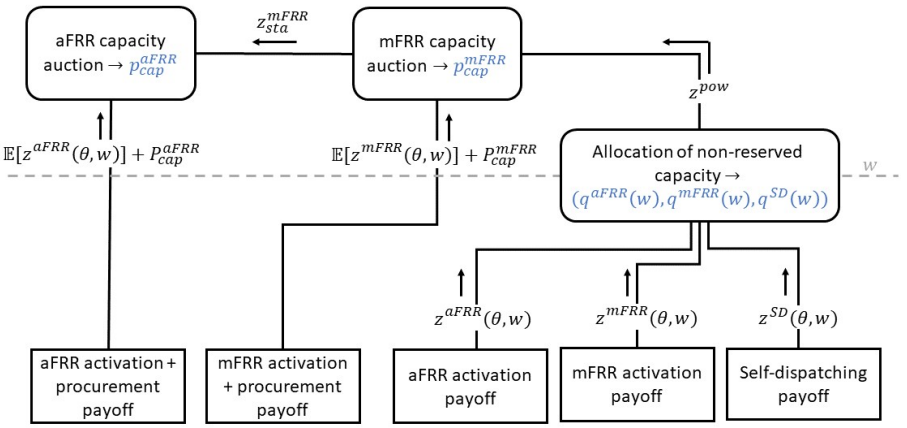


Figure 4.6: Strategy set with balancing capacity market.

This analysis will describe the strategy sets for players participating in the game described in figure 4.5 and will then characterize equilibrium for the corner case with full capacity demand curve, and illustrate it on the example of section 4.4.

4.A.1 Strategy sets with Capacity Markets

The full game with balancing capacity markets is displayed in figure 4.6. The strategy set for an agent with marginal cost θ participating in this game is the concatenation of the strategy set of the initial game for every possible realization of gate-closure forecast (see (4.10) and (4.13)), $(q^{aFRR}(w), q^{mFRR}(w), q^{SC}(w))$, and the aFRR and mFRR capacity price bids, p_{cap}^{aFRR} and p_{cap}^{mFRR}.[14]

$$\mathcal{Q}^F(\theta) = \{(p_{cap}^{aFRR}, p_{cap}^{mFRR}, q^{aFRR}, q^{mFRR}, q^{SD}) | q^{aFRR} + q^{mFRR} + q^{SD} = 1, \\ q^{aFRR} \geq 0, q^{mFRR} \geq 0, q^{SD} \geq 0\} \tag{4.39}$$

The optimal strategy of this game is obtained through backward induction by starting at the bottom of the tree at the aFRR and mFRR activation payoffs and the self-dispatching payoff. The payoff of the non-reserved capacity is then back-propagated to the mFRR capacity auction and then to the aFRR capacity auction.

The energy activation stage, after w is revealed, is not impacted by the balancing capacity auctions and the payoff of the non-reserve capacity, z^{pow},

[14] In practice, agents must submit a price-quantity bid for the mFRR and aFRR balancing capacity auction, $(p_{cap}^{aFRR}, q_{cap}^{aFRR})$ and $(p_{cap}^{mFRR}, q_{cap}^{mFRR})$. We only consider pure strategies between the aFRR and mFRR balancing capacity auctions thus the quantity component of the balancing capacity bid is set at the maximum of the capacity.

can be formulated as the expectation over the realization of w of the maximum of the aFRR and mFRR balancing energy payoff and the self-dispatching payoff.

$$z^{pow} = \int \max\left(z^{aFRR}(\theta, w); z^{mFRR}(\theta, w); z^{SD}(\theta, w)\right) dF_W(w) \tag{4.40}$$

The expected non-reserved payoff is then backpropagated to the mFRR capacity auction (see top-right of figure 4.6). The tradeoff between (i) participating in the mFRR balancing capacity auction and then in the mFRR balancing energy auction and (ii) participating in the balancing market as non-reserved capacity corresponds to the price component of the balancing capacity bid. It determines whether a bid is selected in the capacity auction and is sufficient for characterizing pure strategies. As in the case of non-reserved capacity, bidding the marginal cost in the subsequent balancing energy auction is weakly dominant. The optimal bidding strategy at this stage can then be found by solving (4.41) given an exogenous mFRR capacity price P_{cap}^{mFRR}:

$$\max_{p_{cap}^{mFRR}} \begin{cases} P_{cap}^{mFRR} + \mathbb{E}_W[z^{mFRR}(\theta, w)] & \text{if } p_{cap}^{mFRR} \leq P_{cap}^{mFRR} \\ z^{pow} & \text{if } p_{cap}^{mFRR} > P_{cap}^{mFRR} \end{cases} \tag{4.41}$$

Bidding the opportunity cost of participating in the mFRR balancing energy auction, $z^{pow} - \mathbb{E}_W[z^{mFRR}(\theta, w)]$, is always optimal. If $P_{cap}^{mFRR} + \mathbb{E}_{Y_1}[z^{mFRR}(\theta, w)] > z^{pow}$, then any price bid belonging to the interval $[0, P_{cap}^{mFRR}]$ is optimal. If $P_{cap}^{mFRR} + \mathbb{E}_{Y_1}[z^{mFRR}(\theta, w)] < z^{pow}$, then any price bid belonging to the interval $(P_{cap}^{mFRR}, +\infty)$ is optimal. This optimal strategy results in the expected payoff at the mFRR balancing capacity stage, z_{sta}^{mFRR}, characterized as the maximum between the mFRR procurement and activation payoff and the payoff from not being reserved:

$$z_{sta}^{mFRR} = \max(P_{cap}^{mFRR} + \mathbb{E}_W[z^{mFRR}(\theta, w)]; z^{pow}) \tag{4.42}$$

The expected payoff at the mFRR capacity auction stage is then backpropagated to the aFRR capacity auction (see top-left of figure 4.6). The action space at that stage is similar to that of the mFRR capacity auction stage. The optimal bidding strategy is similar to the one in the mFRR capacity auction except that the tradeoff is between the aFRR procurement and activation payoff, $P_{cap}^{aFRR} + \mathbb{E}_W[z^{aFRR}(\theta, w)]$, and the expected payoff at the mFRR capacity auction stage. This results in the following price offer in the aFRR balancing capacity auction:

$$p_{cap}^{aFRR} = z_{sta}^{mFRR} - \mathbb{E}_W[z^{aFRR}(\theta, w)] \tag{4.43}$$

In summary, rational fringe agents participating in the multi-product reserve and energy balancing auction game (i) allocate their non-reserved capacity between the aFRR and mFRR balancing energy auction, as well as

self-dispatching depending on the highest activation payoff (bottom right of figure 4.6), (ii) bid the difference between solely participating in the mFRR balancing energy auction and participating in the balancing market as non-reserved capacity in the mFRR capacity auction (top right of figure 4.6), and (iii) bid the difference between solely participating in the aFRR balancing energy auction and the payoff at the mFRR capacity auction stage in the aFRR capacity auction (top left of figure 4.6).

4.A.2 Equilibrium with Full Capacity Demand Curve

The equilibria discussed in section 4.3 is a corner case of the complete multi-product reserve and energy games of figure 4.6 with zero aFRR and mFRR capacity demand curves. A second corner case of this game arises when the system operator uses *full capacity demand curves*, which procure the entirety of the balancing capacity for both fast and slow assets in the forward (day-ahead) market. This corresponds to inelastic reserve requirements for aFRR equal to the installed fast capacity and reserve requirements for mFRR equal to the installed slow capacity. This would correspond to aFRR and mFRR capacity demand curve of width N^F and N^S. This scenario results in the following equilibrium.

Proposition 4.7. *The equilibrium with full capacity demand curves is characterized by every fast agent being selected in the aFRR capacity auction and every slow agent being selected in the mFRR capacity auction.*[15]

Proof. At equilibrium, the mFRR capacity price is equal to the difference in payoff between participating in the mFRR balancing energy auction versus participating in the balancing market as non-reserved capacity for the agent with the highest such cost, and the aFRR capacity price is equal to the maximum between (a) the difference between participating in the aFRR balancing energy auction versus participating in the balancing market as non-reserved capacity for the agent with the highest such cost, and (b) the difference between participating in the aFRR balancing energy auction versus participating in the mFRR balancing energy and balancing capacity auctions for the agent with the highest such cost. These opportunity costs are presented in figure 4.7. No price-taking agent has an incentive to deviate from the equilibrium given these capacity prices. □

Full capacity demand curves result in mFRR and aFRR capacity prices that are equal to zero for the "mFRR only" imbalance pricing policy. Participating in the mFRR balancing energy auction is the optimal strategy for slow assets, therefore slow assets have no opportunity cost for doing so, which results in

[15] Technically, this equilibrium is generated from ϵ-*full capacity demand curves* that procure the entirety of the balancing capacity minus a small ϵ. The equilibrium requires a fringe agent to not be selected to ensure that the last agent that is selected in the balancing capacity auction bids its opportunity cost.

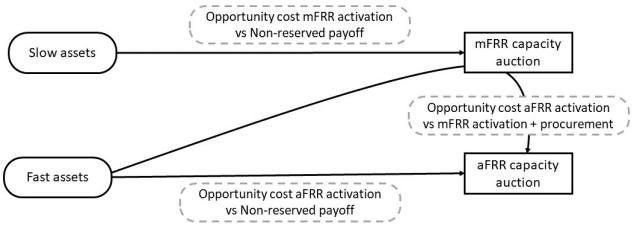


Figure 4.7: Opportunity cost in the capacity auctions.

a zero mFRR capacity price. If there is no mFRR procurement payoff, participating in the aFRR balancing energy auction remains the optimal strategy for fast assets. They too have no opportunity cost, which results in a zero aFRR capacity price. More generally, every balancing energy market design, i.e. the combination of an mFRR activation strategy and an imbalance pricing scheme, that incentivizes agents to participate in the best quality auction they can will result in zero aFRR and mFRR balancing capacity prices. This includes the "mean mFRR and aFRR" imbalance price under the least-cost mFRR activation strategy.

Balancing energy market designs that do not incentivize agents to offer their capacity in the best quality balancing energy auction can generate non-zero balancing capacity prices. Slow assets require a compensation for participating in the mFRR balancing energy auction if it is not their optimal strategy. Even if participating in the aFRR balancing energy auction is still the optimal strategy for the fast non-reserved capacity, the mFRR procurement payoff pushes the aFRR capacity price up as this generates an opportunity cost between solely offering aFRR balancing energy and participating in the mFRR balancing capacity auction (see figure 4.7).

4.A.3 Capacity Prices with Full Capacity Demand Curves for the Illustrative Example

Self-dispatching can be restricted to the case of the "mean mFRR and aFRR" and the "max mFRR and mean aFRR" imbalance pricing schemes given sufficiently large capacity demand curves. The "mFRR only" imbalance price does not incentivize self-dispatching and generates zero capacity prices. This behavior and the resulting mFRR balancing capacity prices are presented in figure 4.8 for the illustrative example. The horizontal axis in this figure represents the slope of a linear mFRR activation strategy up to the least-cost activation strategy at 2/3. The lower the slope, the less mFRR is activated by the TSO, and the higher the opportunity cost for participating in the mFRR auction. Note that the aFRR balancing capacity prices are equal to the mFRR balancing capacity prices as the dominating opportunity cost in figure 4.7 is the one related to participating in the mFRR capacity auction.

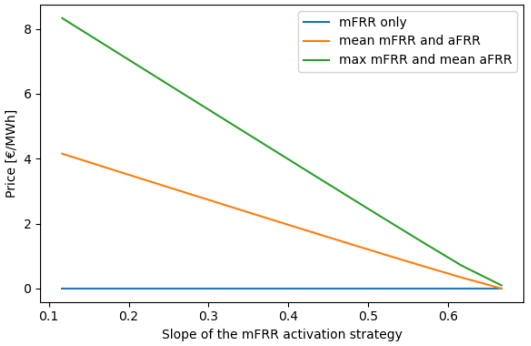


Figure 4.8: mFRR and aFRR capacity prices (which are equal) under full capacity demand curves for an mFRR activation strategy that is lower than or equal to that of the least-cost activation strategy.

5 Conclusion

Market design is the cornerstone of electricity systems for reaching the triple objectives of reliability, affordability and sustainability. It should create an environment that incentivizes the efficient short- and long-run allocation of resources. This thesis discusses the short-run pricing and dispatching of electricity in the context of the integration of European balancing markets and the introduction of scarcity pricing through an ORDC. It is composed of three independent works analysing specific issues related to either scarcity pricing (chapter 2), or the integration of European balancing markets (chapters 4), or both (chapter 3). These works are also characterized by the techniques used. Chapters 3 and 4 rely on game-theoretical techniques to assess different regulatory frameworks concerning the balancing and imbalance prices, whereas chapter 2 uses operations research methods to calibrate or determine demand curves in a given regulation setting.

Chapter 2 provides a tool for assessing the calibration of administrative ORDCs in Belgium. It uses a simulation of short-term electricity markets of Belgium to measure the trade-off between system reliability and cost of operation generated by the ORDC calibration. The simulator is composed of four embedded optimization problems replicating the short-term revelation of uncertainty, the inertia of the power plants, and the decision process behind their commitment and dispatch.

Chapter 3 analyses the introduction of scarcity adders on the balancing and imbalance prices in an integrated European balancing market. It uses a game-theoretical model to show the inefficiencies resulting from the unilateral implementation by a member state of adders without a real-time market for reserve in a cross-border balancing market. The lack of a real-time market for reserve modifies the bidding incentives and generates out-of-merit activations that are detrimental to the member state using the adder because their consumers bear the increased balancing cost.

Chapter 4 discusses the impact of multiple balancing energy products on the balancing process. It uses a game-theoretical model to analytically derive the impact of (i) the mFRR activation strategy, (ii) the imbalance pricing

scheme, and (iii) the balancing capacity demand curve on the balancing process equilibrium. The four main policy insights highlighted by our analysis are (1) the "mFRR only" imbalance price allows for simple optimal strategies and prevents self-scheduling from participants, (2) the minimum balancing activation cost can be reached from the "mFRR only" and "mean mFRR and aFRR" imbalance pricing schemes under the optimal mFRR activation strategy, (3) if the optimal mFRR activation strategy is not available, the "mean mFRR and aFRR" imbalance pricing scheme incentivizes self-scheduling in a way that reduces the balancing cost compared to the no-reaction benchmark, and (4) the "max mFRR and mean aFRR" imbalance price distorts price signals and induces an inefficient level of self-dispatching.

These analyses have also highlighted the need for further research on European balancing markets. For instance, the roll-out of the cross-border balancing platforms and the subsequent occurrences of prices spikes (periods when prices exceed 7500€/MWh) in Austria have demonstrated the need for improved mFRR activation strategies. The use of elastic mFRR demand curves in MARI is a potential method to mitigate those price spikes. Elastic mFRR demand curves would enable the system operator to better manage the trade-off between aFRR and mFRR activation and could allow the operator to get closer to the least-cost mFRR activation strategy discussed in chapter 4. This raises questions about the optimal procedure to determine such elastic demand curves. Parametric cost function approximation is a promising technique for characterizing demand curve [Cartuyvels et al., 2024]. It is a reinforcement learning technique that allows us to model the problem of finding the optimal demand curve parametrization as a stochastic non-convex program. The variables of the problem are the parameters of the demand curve (its width, the number of steps, ...) and the objective is a stochastic cost function. This method could also be applied in other contexts involving administrative demand curves such as an operating reserve demand curve or a capacity demand curve.

Bibliography

[ACER, 2020a] ACER (2020a). Decision No 01/2020 of the European Union Agency for the Cooperation of Energy Regulators of 24 January 2020 on the Methodology to Determine Prices for the Balancing Energy that Results from the Activation of Balancing Energy Bids.

[ACER, 2020b] ACER (2020b). Methodology for the Harmonisation of the main feature of the imbalance settlement in accordance with Article 52(2) of Commission Regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing.

[ACER, 2023] ACER (2023). Progress of EU Electricity Wholesale Market Integration – 2023 Market Monitoring Report. Technical report.

[Atakan et al., 2022] Atakan, S., Gangammanavar, H., and Sen, S. (2022). Towards a sustainable power grid: Stochastic hierarchical planning for high renewable integration. *European Journal of Operational Research*.

[Bakirtzis et al., 2015] Bakirtzis, E. A., Simoglou, C. K., Biskas, P. N., Labridis, D. P., and Bakirtzis, A. G. (2015). Comparison of advanced power system operations models for large-scale renewable integration. *Electric Power Systems Research*, 128:90–99.

[Boiteux, 1960] Boiteux, M. (1960). Peak-Load Pricing. *The Journal of Business*, 33(2):157–179. Publisher: University of Chicago Press.

[Borenstein et al., 2002] Borenstein, S., Bushnell, J. B., and Wolak, F. A. (2002). Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market. *American Economic Review*, 92(5):1376–1405.

[Bottieau et al., 2021] Bottieau, J., Bruninx, K., Sanjab, A., De Grève, Z., Vallée, F., and Toubeau, J.-F. (2021). Automatic risk adjustment for profit maximization in renewable dominated short-term electricity markets. *International Transactions on Electrical Energy Systems*, 31(12):e13152. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2050-7038.13152.

[Brijs et al., 2017] Brijs, T., De Jonghe, C., Hobbs, B. F., and Belmans, R. (2017). Interactions between the design of short-term electricity markets in the CWE region and power system flexibility. *Applied Energy*, 195:36–51.

[Brown, 2018] Brown, D. P. (2018). Capacity payment mechanisms and investment incentives in restructured electricity markets. *Energy Economics*, 74:131–142.

[Bushnell et al., 2017] Bushnell, J. B., Holland, S. P., Hughes, J. E., and Knittel, C. R. (2017). Strategic Policy Choice in State-Level Regulation: The EPA's Clean Power Plan. *American Economic Journal: Economic Policy*, 9(2):57–90.

[Bushnell and Oren, 1994] Bushnell, J. B. and Oren, S. S. (1994). Bidder cost revelation in electric power auctions. *Journal of Regulatory Economics*, 6(1):5–26.

[Cartuyvels et al., 2023] Cartuyvels, J., Bertrand, G., and Papavasiliou, A. (2023). Market Equilibria in Cross-Border Balancing Platforms. *IEEE Transaction on Energy Markets, Policy and Regulation*. under review.

[Cartuyvels et al., 2024] Cartuyvels, J., Bertrand, G., and Papavasiliou, A. (2024). Efficient Dispatch in Cross-Border Balancing Platforms: Elastic Demand Through Parametric Cost Function Approximation. In *2024 20th International Conference on the European Energy Market (EEM)*, Istanbul, Turkyie.

[Cartuyvels and Papavasiliou, 2023] Cartuyvels, J. and Papavasiliou, A. (2023). Calibration of Operating Reserve Demand Curves Using a System Operation Simulator. *IEEE Transactions on Power Systems*, 38(4):3043–3055.

[Cervigni and Perekhodtsev, 2013] Cervigni, G. and Perekhodtsev, D. (2013). Wholesale electricity markets. *Chapters*, pages 18–66. Publisher: Edward Elgar Publishing.

[Chao and Wilson, 2002] Chao, H.-P. and Wilson, R. (2002). Multi-Dimensional Procurement Auctions for Power Reserves: Robust Incentive-Compatible Scoring and Settlement Rules. *Journal of Regulatory Economics*, 22(2):161–183.

[Chaves-Ávila et al., 2014] Chaves-Ávila, J. P., van der Veen, R. A. C., and Hakvoort, R. A. (2014). The interplay between imbalance pricing mechanisms and network congestions – Analysis of the German electricity market. *Utilities Policy*, 28:52–61.

[Cho and Papavasiliou, 2024] Cho, J. and Papavasiliou, A. (2024). Exact Mixed-Integer Programming Approach for Chance-Constrained Multi-Area Reserve Sizing. *IEEE Transactions on Power Systems*, 39(2):3310–3323. Conference Name: IEEE Transactions on Power Systems.

[Commision, 2019] Commision, E. (2019). Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast) (Text with EEA relevance.). Legislative Body: EP, CONSIL.

[CONSENTEC, 2022] CONSENTEC (2022). Description of the Balancing Process and the Balancing Markets in Germany. Technical report.

[Cramton, 2017] Cramton, P. (2017). Electricity market design. *Oxford Review of Economic Policy*, 33(4):589–612.

[Cramton and Stoft, 2005] Cramton, P. and Stoft, S. (2005). A Capacity Market that Makes Sense. *The Electricity Journal*, 18(7):43–54.

[CREG, 2018] CREG (2018). Décision 1806. Technical report, Commission for Electricity and Gas Regulation, Brussels.

[CREG, 2021] CREG (2021). Study on the implementaion of a scarcity pricing mechanism in Belgium. Technical Report F2203, Commission for Electricity and Gas Regulation, Brussels.

[CREG, 2023] CREG (2023). Projet de décision sur la révision des méthodologies et des conditions pour le responsable d'équilibre ou « les T&C BRP » par la CREG dans le cadre de l'intégration du calcul du prix de déséquilibre. Technical report.

[Daraeepour et al., 2019] Daraeepour, A., Patino-Echeverri, D., and Conejo, A. J. (2019). Economic and environmental implications of different approaches to hedge against wind production uncertainty in two-settlement electricity markets: A PJM case study. *Energy Economics*, 80:336–354.

[Department for Business, 2020] Department for Business, E. a. I. S. (2020). GB Implementation Plan. Technical report, Department for Business, Energy and Industrial Strategy, London.

[Devogelaer, 2017] Devogelaer, D. (2017). Increasing interconnections: to build or not to build, that is (one of) the question(s). Technical report, Federal Planning Bureau, Brussels.

[E-Bridge Consulting, 2014] E-Bridge Consulting (2014). Potential Cross-Border Balancing Cooperation Between the Belgian, Dutch and German Electricity Transimission System Operators. Technical report, 50Hertz Transmission GmbH, Amprion GmbH, Elia System Operator NV, TenneT TSO B.V., TenneT TSO GmbH, TransnetBW GmbH, Bonn, Germany.

[Ehrhart and Ocker, 2021] Ehrhart, K.-M. and Ocker, F. (2021). Design and regulation of balancing power auctions: an integrated market model approach. *Journal of Regulatory Economics*, 60(1):55–73.

[EirGrid, 2017] EirGrid (2017). Training Modules, Chapter 10: Administered Scarcity Pricing and Reserve Scarcity Pricing.

[ELIA, 2018] ELIA (2018). Study report on Scarcity Pricing in the context of the 2018 discretionary incentives. Technical report, ELIA Group, Brussels.

[ELIA, 2019] ELIA (2019). Adequacy and flexibility study for Belgium 2020-2030. Technical report, ELIA, Brussels.

[ELIA, 2021a] ELIA (2021a). Data Download Page.

[ELIA, 2021b] ELIA (2021b). Proposal for Modification of the Tariff for Maintaining and Restoring the Residual Balance of Individual Access Responsible Parties. Technical report, ELIA, Brussels.

[ELIA, 2022] ELIA (2022). Tarifs pour le Maintien et la Restauration de l'Equilibre Résiduel des Responsables d'Accès Individuel, Période 2020-2023. Technical report, Brussels.

[ELIA, 2023] ELIA (2023). ELIA answer to CREG public consultation from 05/01 to 06/02 on draft decision (B)2947 regarding modifications in the T&C BRP by the CREG. Technical report.

[ELIA, 2024] ELIA (2024). Adequacy and Flexibility Study for Belgium 2024-2034. Technical report.

[Fabra et al., 2006] Fabra, N., von der Fehr, N.-H., and Harbord, D. (2006). Designing Electricity Auctions. *The RAND Journal of Economics*, 37(1):23–46. Publisher: [RAND Corporation, Wiley].

[Farahmand and Doorman, 2012] Farahmand, H. and Doorman, G. L. (2012). Balancing market integration in the Northern European continent. *Applied Energy*, 96:316–326.

[Gorman, 2022] Gorman, W. (2022). The quest to quantify the value of lost load: A critical review of the economics of power outages. *The Electricity Journal*, 35(8):107187.

[Hellmers et al., 2016] Hellmers, A., Zugno, M., Skajaa, A., and Morales, J. M. (2016). Operational Strategies for a Portfolio of Wind Farms and CHP Plants in a Two-Price Balancing Market. *IEEE Transactions on Power Systems*, 31(3):2182–2191.

[Hogan, 2016] Hogan, W. (2016). Electricity Market Design: Optimization and Market Equilibrium.

[Hogan, 2005] Hogan, W. W. (2005). On an "Energy Only" electricity market design for resource adequacy. White Paper, Center for Business and Government, JFK School of Government, Harvard University, Cambridge, MA.

[Hogan, 2013] Hogan, W. W. (2013). Electricity Scarcity Pricing Through Operating Reserves. *Economics of Energy & Environmental Policy*, Volume 2(Number 2).

[Hogan, 2019] Hogan, W. W. (2019). Market Design Practices: Which Ones Are Best? [In My View]. *IEEE Power and Energy Magazine*, 17(1):100–104.

[Hogan and Pope, 2019] Hogan, W. W. and Pope, S. L. (2019). PJM Reserve Markets: Operating Reserve Demand Curve Enhancements. Technical report, Center for Business and Government, JFK School of Government, Harvard University, Cambridge, MA.

[Hogan and Pope, 2020] Hogan, W. W. and Pope, S. L. (2020). Real-Time Co-Optimization with the ERCOT ORDC. Technical report, Center for Business and Government, JFK School of Government, Harvard University, Cambridge, MA.

[IPCC, 2023] IPCC (2023). Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. Technical report, Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland. Please access report for extended list of writing team, editors, scientific steering committee and other contributing authors.

[Jevons, 1871] Jevons, W. S. (1871). The Theory of Political Economy. *History of Economic Thought Books*. Publisher: McMaster University Archive for the History of Economic Thought.

[Joskow and Tirole, 2007] Joskow, P. and Tirole, J. (2007). Reliability and competitive electricity markets. *The RAND Journal of Economics*, 38(1):60–84. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-2171.2007.tb00044.x.

[Joskow, 2008] Joskow, P. L. (2008). Capacity payments in imperfect electricity markets: Need and design. *Utilities Policy*, 16(3):159–170.

[Joskow, 2019] Joskow, P. L. (2019). Challenges for wholesale electricity markets with intermittent renewable generation at scale: the US experience. *Oxford Review of Economic Policy*, 35(2):291–331.

[Kamat and Oren, 2002] Kamat, R. and Oren, S. S. (2002). Rational Buyer Meets Rational Seller: Reserves Market Equilibria under Alternative Auction Designs. *Journal of Regulatory Economics*, 21(3):247–288.

[Kirschen and Strbac, 2018] Kirschen, D. S. and Strbac, G. (2018). *Fundamentals of Power System Economics*. John Wiley & Sons. Google-Books-ID: I9hhDwAAQBAJ.

[Lavin et al., 2020] Lavin, L., Murphy, S., Sergi, B., and Apt, J. (2020). Dynamic operating reserve procurement improves scarcity pricing in PJM. *Energy Policy*, 147:111857.

[Mas-Colell et al., 1995] Mas-Colell, A., Whinston, M. D., Green, J. R., Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press, Oxford, New York.

[Matsumoto et al., 2021] Matsumoto, T., Bunn, D. W., and Yamada, Y. (2021). Mitigation of the Inefficiency in Imbalance Settlement Designs using Day-Ahead Prices. *IEEE Transactions on Power Systems*, pages 1–1. Conference Name: IEEE Transactions on Power Systems.

[Mays, 2021] Mays, J. (2021). Quasi-Stochastic Electricity Markets. *INFORMS Journal on Optimization*, 3(4):350–372. Publisher: INFORMS.

[Mays and Jenkins, 2023] Mays, J. and Jenkins, J. D. (2023). Financial Risk and Resource Adequacy in Markets With High Renewable Penetration. *Policy and Regulation IEEE Transactions on Energy Markets*, 1(4):523–535. Conference Name: Policy and Regulation IEEE Transactions on Energy Markets.

[Meeus, 2020] Meeus, L. (2020). *The Evolution of Electricity Markets in Europe*. Loyola De Palacio on European Energy Policy. Edward Elgar Publishin.

[Nederlands, 2023] Nederlands, N. (2023). Codewijzigingsvoorstel implementatie PICASSO en MARI.

[Newbery et al., 2018] Newbery, D., Pollitt, M. G., Ritz, R. A., and Strielkowski, W. (2018). Market design for a high-renewables European electricity system. *Renewable and Sustainable Energy Reviews*, 91:695–707.

[Newbery et al., 2016] Newbery, D., Strbac, G., and Viehoff, I. (2016). The benefits of integrating European electricity markets. *Energy Policy*, 94:253–263.

[NYISO, 2019] NYISO (2019). Ancillary Services Shortage Pricing. Technical report, New York ISO, NY.

[Ocker et al., 2018] Ocker, F., Ehrhart, K.-M., and Belica, M. (2018). Harmonization of the European balancing power auction: A game-theoretical and empirical investigation. *Energy Economics*, 73:194–211.

[Oren, 2001] Oren, S. (2001). Design of ancillary service markets. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pages 9 pp.–.

[Papavasiliou, 2020] Papavasiliou, A. (2020). Scarcity pricing and the missing European market for real-time reserve capacity. *The Electricity Journal*, 33(10):106863.

[Papavasiliou and Avila, 2024] Papavasiliou, A. and Avila, D. (2024). Welfare Benefits of Co-Optimizing Energy and Reserves. Technical report, ACER.

[Papavasiliou and Bertrand, 2021] Papavasiliou, A. and Bertrand, G. (2021). Market Design Options for Scarcity Pricing in European Balancing Markets. *IEEE Transactions on Power Systems*, 36(5):4410–4419.

[Papavasiliou et al., 2023] Papavasiliou, A., Cartuyvels, J., Bertrand, G., and Marien, A. (2023). Implementation of scarcity pricing without co-optimization in European energy-only balancing markets. *Utilities Policy*, 81:101488.

[Papavasiliou and Smeers, 2017] Papavasiliou, A. and Smeers, Y. (2017). Remuneration of Flexibility using Operating Reserve Demand Curves: A Case Study of Belgium. *The Energy Journal*, 38(01).

[Papavasiliou et al., 2018] Papavasiliou, A., Smeers, Y., and Bertrand, G. (2018). An Extended Analysis on the Remuneration of Capacity under Scarcity Conditions. *Economics of Energy & Environmental Policy*, 7(2).

[Papavasiliou et al., 2019] Papavasiliou, A., Smeers, Y., and de Maere d'Aertrycke, G. (2019). Study on the general design of a mechanism for the remuneration of reserves in scarcity situations. Technical report, Center for Operations Research and Econometrics, UCLouvain, Louvain-la-Neuve, Belgium.

[Papavasiliou et al., 2021] Papavasiliou, A., Smeers, Y., and de Maere d'Aertrycke, G. (2021). Market Design Considerations for Scarcity Pricing: A Stochastic Equilibrium Framework. *The Energy Journal*, 42(01).

[Petitet et al., 2019] Petitet, M., Perrot, M., Mathieu, S., Ernst, D., and Phulpin, Y. (2019). Impact of gate closure time on the efficiency of power systems balancing. *Energy Policy*, 129:562–573.

[Poplavskaya et al., 2020] Poplavskaya, K., Lago, J., and de Vries, L. (2020). Effect of market design on strategic bidding behavior: Model-based analysis of European electricity balancing markets. *Applied Energy*, 270:115130.

[Poplavskaya et al., 2021] Poplavskaya, K., Lago, J., Strömer, S., and de Vries, L. (2021). Making the most of short-term flexibility in the balancing market: Opportunities and challenges of voluntary bids in the new balancing market design. *Energy Policy*, 158:112522.

[Prabhakar Karthikeyan et al., 2013] Prabhakar Karthikeyan, S., Jacob Raglend, I., and Kothari, D. P. (2013). A review on market power in deregulated electricity market. *International Journal of Electrical Power & Energy Systems*, 48:139–147.

[SEM, 2021] SEM (2021). Discussion Paper and Call for Evidence on Scarcity Pricing and Demand Response in the SEM. Technical report, Single Electricity Market System Operator, Dublin.

[Shinde et al., 2021] Shinde, P., Hesamzadeh, M. R., Date, P., and Bunn, D. W. (2021). Optimal Dispatch in a Balancing Market With Intermittent Renewable Generation. *IEEE Transactions on Power Systems*, 36(2):865–878.

[Simao et al., 2017] Simao, H., Powell, W., Archer, C., and Kempton, W. (2017). The challenge of integrating offshore wind power in the U.S. electric grid. Part II: Simulation of electricity market operations. *Renewable Energy*, 103:418–431.

[Simoglou et al., 2010] Simoglou, C. K., Biskas, P. N., and Batirtizis, A. G. (2010). Optimal Self-Scheduling of a Thermal Producer in Short-Term Electricity Market by MILP. *IEEE Transactions on Power Systems*, 25(4).

[Smets et al., 2023] Smets, R., Bruninx, K., Bottieau, J., Toubeau, J.-F., and Delarue, E. (2023). Strategic Implicit Balancing with Energy Storage Systems via Stochastic Model Predictive Control. *Policy and Regulation IEEE Transactions on Energy Markets*, pages 1–14.

[Stattnett et al., 2023] Stattnett, Energinet, fraftnat, S., and Fingrid (2023). Option Space for Future Imbalance Pricing in the Nordics once Connected to the European Balancing Energy Platforms. Technical report.

[Stoft, 2002] Stoft, S. (2002). *Power System Economics: Designing Markets for Electricity*. IEEE Press Wiley, Piscataway, NJ.

[van der Veen et al., 2012] van der Veen, R. A. C., Abbasy, A., and Hakvoort, R. A. (2012). Agent-based analysis of the impact of the imbalance pricing mechanism on market behavior in electricity balancing markets. *Energy Economics*, 34(4):874–881.

[Wilson, 2002] Wilson, R. (2002). Architecture of Power Markets. *Econometrica*, 70(4):1299–1340.

[Zarnikau et al., 2020] Zarnikau, J., Zhu, S., Woo, C. K., and Tsai, C. H. (2020). Texas's operating reserve demand curve's generation investment incentive. *Energy Policy*, 137(C). Publisher: Elsevier.

[Zhao et al., 2018] Zhao, F., Zheng, T., and Litvinov, E. (2018). Constructing Demand Curves in Forward Capacity Market. *IEEE Transactions on Power Systems*, 33(1):525–535. Conference Name: IEEE Transactions on Power Systems.

[Zhou and Botterud, 2014] Zhou, Z. and Botterud, A. (2014). Dynamic Scheduling of Operating Reserves in Co-Optimized Electricity Markets With Wind Power. *IEEE Transactions on Power Systems*, 29(1):160–171. Conference Name: IEEE Transactions on Power Systems.