NICOLAS STEVENS

PRICE FORMATION WITH NON-CONVEXITIES: THEORY AND APPLICATIONS FOR THE ELECTRICITY MARKET

Université catholique de Louvain
École Polytechnique de Louvain

Center for Operations Research and Econometrics

Price Formation with Non-Convexities: Theory and Applications for the Electricity Market

Nicolas Stevens

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)
Bert Willems (UCLouvain, Belgium)

Jury:
Mette Bjørndal (NHH, Norway)
Philippe Chevalier (UCLouvain, Belgium)
Richard O'Neill (ARPA-E, United States of America)
Yves Smeers (UCLouvain, Belgium)

Chair:
Philippe Chevalier (UCLouvain, Belgium)

PHD ORGANIZATION

Nicolas Stevens
UCLouvain
École Polytechnique de Louvain
Center for Operations Research and Econometrics

THESIS SUPERVISORS

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

Bert Willems
Professor, UCLouvain
Louvain School of Management
Center for Operations Research and Econometrics

SUPERVISORY COMMITTEE

Mette Bjørndal
Professor, Norwegian School of Economics (NHH)
Department of Business and Management Science

Alexandre Street
Professor, Pontifical Catholic University of Rio de Janeiro
Department of Electrical Engineering

"The old problem asks which unit should be committed; the new problem asks what market design will best solve the old problem."

(Stoft, 2002)

ABSTRACT

Since the restructuring policies that led to the liberalization of the electricity sector and to the existence of a market for power, electricity markets have been organised in a highly centralized fashion, relying on uniform-price auctions. These auctions typically include non-convex bids. The main implication of these non-convexities is that they impede the existence of a uniform "market-clearing" price. Therefore, what the electricity price should be under these settings is an open question, which has attracted the interest of both academics and practitioners over the past thirty years. This dissertation is composed of three main essays that study various aspects of this question.

The first part of the dissertation studies some *short-term* impacts of the market failure we are interested in, namely the non-convexities in the production processes. We look first at the economic properties of various pricing solutions that have been envisioned in the literature, or implemented by some auctioneers. We analyse six different pricing methods and we establish several mathematical properties for enabling their accurate comparison. The findings are illustrated on stylized examples and numerical simulations that are performed on realistic auction datasets. Both theoretical and numerical evidences that are gathered point towards the advantages of the so-called "convex hull pricing" method. The dissertation then goes on with the analysis of some computational challenges. If convex hull pricing is proven to come with several advantages, it is also known for being computationally difficult to calculate. In this chapter, we propose a dual decomposition algorithm known as the "level method" and we adapt the basic algorithm to the specificities of convex hull pricing. We provide empirical evidence about the favorable performance of our algorithm on large test instances.

The second part of the dissertation studies some *long-term* impacts of non-convexities. If the topic of pricing non-convexities in power markets has been primarily focused on indivisibilities in short-term auctions, this second part analyses a source of non-convexities that is not discussed as broadly: the indivisibilities in investment decisions. The absence of equilibrium that we are concerned about is the *long-term* equilibrium. We derive a capacity expansion model with indivisibilities and we highlight the failure arising from it. We then investigate to what extent a capacity market can mitigate the lumpiness problem. We illustrate the main findings with a numerical experiment conducted on the capacity expansion model used by ENTSO-E to assess the adequacy of the European system.

ACKNOWLEDGEMENTS

Four years of PhD have left me with huge intellectual debts, and these acknowledgements are meant to pay my bill to the people that made this journey possible. Scientific research is rarely a lonely and heroic quest of truth, although this idea admittedly sounds romantic.

More than anybody, this thesis owes much to Anthony Papavasiliou. Beyond his significant academic contribution to the content of the three main essays that are the substance of this dissertation, Anthony has also been a very inspiring figure, who has profoundly influenced my own character for almost a decade. His intellectual honesty, his modesty, his rigour in the conduct of scientific research, are among the virtues that are I believe essential to a good scientist, and for which he has been a formidable instructor, which—as it is often the case in moral instruction—means a model that educates by his example. Most valuable for a PhD student, Anthony also demonstrates a boundless availability for his students despite his busy agenda—he truly cares, which is a rare quality. He has also consistently offered me with the most complete freedom to choose the research I wanted to carry out, which I view as a great privilege of scientific work.

Two other faculties have played an important supervising role during my PhD. Yves Smeers, besides his involvement on the chapters 3 and 5 of this dissertation, has become almost like a second supervisor. His wise and always accurate comments, his inexhaustible ability to come up with new ideas, and—mostly—his scepticism in front of any piece of work that merely *asserts* propositions instead of rigorously *defending* them, have greatly contributed to my academic education. Bert Willems has generously accepted to endorse the role of co-supervisor for my PhD. His presence at the Center for Operations Research and Econometrics since 2023 has been particularly stimulating for research. His view, as an economist, on my dissertation was most welcome and has been determinant for my own thinking about future research.

Moreover, I would also like to warmly thank Philippe Chevalier, Mette Bjørndal and Richard O'Neill for being part of the jury of this thesis, as well as for the fruitful discussions during the private defense, and the numerous comments they have provided. My gratitude also goes to Mathieu Van Vyve and Alexandre Street, who have been members of the supervisory committee of this thesis, and have provided valuable comments during the four years of the PhD, particularly during the PhD confirmation.

Each of the three published papers, that led to the chapters 3, 4 and 5 of this dissertation, have greatly benefited from valuable comments and feedbacks from a number of people that I am in debt with. Chapter 3 benefited from discussions

and comments from Mehdi Madani, Richard O'Neill and Alain Marien. Back in 2016, during my master thesis, it was Yurii Nesterov who initially guided me towards the Level Method, the cornerstone of chapter 4. This chapter also benefited from the advises from Gauthier de Maere d'Aertrycke. Fruitful comments were provided by Mehdi Madani, William Hogan, Panagiotis Andrianesis, Ross Baldick and Alejandro Angulo. Daniel Avila provided me with an outstanding help for the processing of the data as well as for running the numerical simulations of chapter 5. Richard O'Neill, Benjamin Hobbs, Jens Weibezahn have also provided remarks on the first version of this chapter.

As the reader is about to find out, throughout this dissertation as often in scientific research, I study a set of *very specific* questions. However, one of the finest pleasure I have in doing my research is that it gives me an excuse for learning a great deal about broader disciplines, and to read what many distinguished scientists have produced in their own research. In this perspective, UCLouvain and the Center for Operations Research and Econometrics have proven to be a stimulating learning environment, thanks to the faculties and PhD fellows working there. Hence, I would like to thank Jacques Cartuyvels, Antoine Germain, Martial Toniotti, Quentin Lété, Thomas De Munck, Jehum Cho, Daniel Avila, Gilles Bertrand, Hakimeh Hemmatipour, Henri Dehaybe, Thomas Eisfeld, Ilyes Mezghani, Céline Gérard, Lucie Paré, Brieuc Pierre, Gianmarco Luu and all the other PhD fellows, faculties and administrative staff for contributing to this learning environment. I would like to extend my special thank to François Maniquet and Amma Panin for teaching to an engineer some of the economic wisdom, as well as to Gregory Ponthière for our frequent and valuable discussions during the last two years, and to Gauthier de Maere with whom I learned a lot by being a teaching assistant for his class. Finally, my PhD has greatly benefited from a research visit at the Harvard Kennedy School in fall 2024. This was possible thanks to William Hogan, whose scientific work has had a profound influence on my own research—in fact, his work turns out to be the foundational idea of the entire dissertation. I am grateful for our various conversations at Harvard which have been a source of inspiration for me and my future research.

At last, I want to turn to some more intimate contributors. For friendship, at its best, is a matter of shared value, mutual education and emulation, the fine relationship I share with many dear friends have shaped my character throughout the many years I have spent in their good company. My whole family, along with my in-laws, make up one of the finest society I know for a researcher to grow up. In particular, and besides being the mere condition of possibility of my existence, my parents have done much, through their education and by their example, to encourage my appetite for knowledge and my intellectual curiosity. Eugénie has been my beloved partner for a decade now. She has made me better in many ways, and has contributed to this PhD journey more than I could tell here.

The PhD was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 850540) as well as by the FEVER Project (grant agreement No. 864537). The research visit at Harvard, in fall 2024, was funded by the Wallonia-Brussels International (WBI) Excellence 2024-2025 grant.

CONTENTS

1

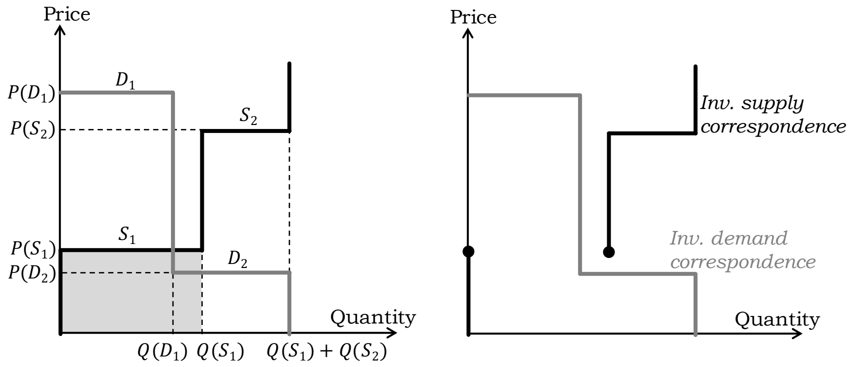
INTRODUCTION

This thesis is composed of three main essays, all of them having in common the study of one fundamental assumption across economics, namely the *convexity* of the economy. In the idealized world of microeconomic theory, consumers hold well-behaved convex preferences over convex consumption sets, and the production processes of firms are described by convex production sets. These assumptions are certainly convenient for deriving a set of remarkable results of great generality. However, they also fall short in accounting for certain aspects of reality. This is obviously known by economists. For example, Debreu explicitly emphasizes that "the inclusion of indivisible commodities" is one of the "important and difficult questions [that] are not answered by the approach taken" in his *Theory of Value* (Debreu, 1959, p. 36). Discussing the convexity of consumer's preferences, Starr ironically comments that "one may be indifferent between an automobile and a boat, but in most cases one can neither drive nor sail the combination of half boat, half car" (Starr, 1969).

Electricity markets turn out to be a domain of economic activity where the defect of the convexity assumption has been particularly apparent. Since the deregulation policies that led to the liberalization of power systems and to the existence of a market for power, electricity markets have typically been organised in a highly centralized fashion, relying on sealed-bid uniform-price auctions. Most of these auctions—in particular the one held in the day ahead—in the US as well as in Europe or in India, include non-convex bids. Therefore, the rules and principles that should be adopted by the auctioneer to clear the non-convex market and to compute prices are not merely a theoretical issue but a very practical one. This pressing need to understand the relationship between the non-convexities in the market and the possible price formation rules has put the topic under the scrutiny of electricity economists and power system engineers, feeding a vivid field of research for the past twenty years.

STYLIZED EXAMPLE OF THE PROBLEM. The main challenge stemming from non-convexities can be efficiently conveyed by examining the example of Figure 1.1. Figure 1.1a presents a hypothetical elementary market with two two demand bids, D_1 and D_2, as well as two supply offers S_1 and S_2. If all these bids were convex, the surplus-maximizing allocation would be to clear D_1, to partially clear S_1 up to $Q(D_1)$ and to reject the other bids. Under this configuration, the market surplus is the area between bids D_1 and S_1, and the price $\pi = P(S_1)$ clears the market: this price, together with the aforementioned allocation, is a competitive equilibrium. The subject matter of this thesis concerns what happens when, say S_1, in this example, is non-convex. If S_1 is an "all-or-nothing" offer (either it is entirely cleared by the auctioneer, or entirely rejected), then the surplus-maximizing allocation is to clear S_1, D_1, as well as a share of D_2 ($Q(S_1) - Q(D_1)$). The surplus is the area between bids D_1 and S_1, *minus* the area between bids S_1 and D_2. As we observe, in this example, the introduction of the indivisibility constraint has only a mild effect on the total achievable surplus. But the issue is that there exists no price that supports this allocation: *there is no competitive equilibrium in this non-convex market*.

A straightforward way to view this is Figure 1.1b. This figure draws the supply and demand correspondences of the market of Figure 1.1a with the non-convex bid S_1. As one may observe, these supply and demand curves *do not intersect*. Another way to explain the inexistence of a competitive equilibrium, a way that is useful when apprehending this problem as a "two-step" process in which the auctioneer first selects the bids that are cleared and then computes a price, is the following. Assuming the surplus-maximizing allocation is selected, the auctioneer has to find a price that is acceptable for all the market participants. If $\pi = P(D_2)$, then the demand bids D_1 and D_2, as well as the supply bid S_2, would have the incentive to implement the cleared allocation. But the supply bid S_1 has an incentive to deviate: this supplier is losing money for each MWh that he produces under this price. One may instead pick a price $\pi = P(S_1)$. Under this choice, S_1 has the incentive to implement the cleared allocation. But the demand bid D_2 now has incentives to deviate from the allocation instructed by the auctioneer, as his willingness-to-pay for electricity consumption is below the market price. Clearly, any price in the interval $[P(D_2), P(S_1)]$



(a) Supply & demand bids.

(b) Supply & demand correspondences.

FIGURE 1.1: Stylized electricity market with two demand bids (D_1 and D_2) and two supply bids (S_1 and S_2) with one non-convexity: bid S_1 is an "all-or-nothing" offer.

would lead to a similar issue and be unsatisfactory for some of the market participants, and the same is obviously true for prices out of this interval.

This type of analysis led Herbert Scarf to the following diagnosis:

> "[...] in the presence of indivisibilities in production, prices simply don't do the jobs that they were meant to do." (Scarf, 1994)

As far as electricity markets are concerned, this reasoning leaves us with one question—*at the end of the day, what should be the right price of electricity under the settings of this example?* The rest of the thesis studies several facets of this question.

ORGANISATION OF THE MATERIAL. The material of the thesis is organised as follows. Chapter 2 outlines the main concepts and building blocks of an electricity market, and explains the paramount institutional arrangements as we know them in Europe or in the US. Understanding these arrangements is important for the next chapters of the thesis. Electricity markets are indeed highly sophisticated *institutions*: trades do not take place on an invisible "backstage", but happen on regulated marketplaces which have been designed for decades. As Hogan (2002) phrases it, "*power markets are made, they don't just happen*." To say it differently, electricity markets are organised in such a way that the "hand of the market" is rather *visible*. The pricing principles that are going to be studied in the thesis are not just theoretical objects that emulate what happens on the backstage, but should be viewed as various possible "rules of the game" that could be implemented in the actual organisation of the market.

The rest of the thesis studies price formation in presence of non-convexities in the supply curve, as introduced above. According to Cramton (2017), the implementation of a market for power had two main objectives: short-term efficiency and long-term efficiency. The material of the thesis follows this structure. The first part of the thesis studies some *short-term* impacts of the market failure we are interested in, namely the non-convexities in the production processes. This first part is further divided into chapters 3 and 4. The second part of the thesis studies some *long-term* impacts of non-convexities. This corresponds to chapter 5.

Chapter 3 goes on with the analysis of multiple price formation rules that have been advocated for electricity auctions in the presence of non-convexities[1]. Although marginal pricing has traditionaly been contemplated as the Holy Grail of economic theory, its usage has occasionally been challenged, at least since Coase (1946) discussed it in "the Marginal Cost Controversy", which studies a situation of increasing return to scale (see also Coase (1970)). Coase's analysis suggests that the defect of marginal pricing in this situation could be solved by introducing some sort of discrimination or "multi-part pricing", as he names it. In spirit, this is similar to what has been proposed in the past two decades for power auctions. Returning to the example of Figure 1.1, the solution of Coase would imply pricing electrical energy at the marginal cost of the cleared allocation. In the example, this is $\pi = P(D_2)$. Then, as supplier S_1 is facing a loss of $(P(S_1) - \pi)Q(S_1)$, the market operator would provide this supplier with a discriminatory payment covering his losses. This multi-part pricing restores the equilibrium: the market clears and each agent in the market has the incentive to implement the cleared allocation. Nonetheless, one may further point out that an alternative, in the same spirit, could be $\pi = P(S_1)$, together with some—lesser—discriminatory payments to consumer D_2. Many other alternatives exist. Which one should be preferred? Chapter 3 reviews these alternatives and formalizes them on a general model of an electricity market. Several mathematical properties of these pricing schemes are established and illustrated numerically on auction datasets of realistic size.

Chapter 4 proceeds with the analysis of some computational challenges related to the pricing of non-convexities in electricity auctions[2]. One feature of the analysis of electricity markets—which, in my opinion, also

1 Chapter 3 reproduces the text, with minor changes, of the following published paper: Nicolas Stevens, Anthony Papavasiliou and Yves Smeers. "On some advantages of convex hull pricing for the European electricity auction." *Energy Economics* 134 (2024): 107542.

2 Chapter 4 reproduces the text, with minor changes, of the following published paper:

renders the field exciting and enjoyable—is that it lies at the intersection of several disciplines, such as economics and power system engineering, mathematical modelling and mathematical programming. The topic of pricing non-convexities is no exception. Chapter 3 argues for some advantages of one particular pricing approach, the so-called "convex hull pricing" scheme. The principle of this approach is to compute the uniform price of energy that minimizes the amount of discriminatory payments that are implied. In the example of Figure 1.1, the computation of the convex hull price is straightforward: it corresponds to $\pi = P(S_1)$. The approach boils down to computing the convex hull of suppliers' production sets which, in the case of Figure 1.1, is immediate as it corresponds to its linear relaxation. Whereas this is direct in our stylized example, in general, it is a computationally challenging problem. Chapter 4 reviews several approaches that have been proposed in the scientific literature and then develops an algorithm, the so-called "level method", to compute locational convex hull prices. The algorithmic procedure is described and implemented on several instances of auctions of realistic size.

With chapter 5, we move to the second part of the thesis that studies long-term implications of non-convexities.[3] Chapter 3 and 4 focus with the functioning of short-term electricity auctions, considering both economic and algorithmic stakes. Chapter 5 is concerned with an investment problem in which the non-convexities stem from the indivisible nature of investment decisions. The chapter builds on the seminal paper of Scarf (1994), who highlighted some of the issues caused by the presence of indivisibilities in investment decisions[4]. Because of economies of scale in production processes, the electricity sector has been characterised by the promotion of large and fundamentally indivisible assets. Since investment comes with such large lumps of capacity, the optimal investment choices might lead to slight "over-capacity" which, if economically efficient, may also in turn pull down the electricity price, rendering the investment unprofitable in the first place, thus preventing the entry of these capacities. In spirit, this is similar to the case of Figure 1.1, where clearing supplier S_1 leads to a marginal price that does not cover production costs. Chapter 5 analyses

---

Nicolas Stevens and Anthony Papavasiliou. "Application of the level method for computing locational convex hull prices." *IEEE Transactions on Power Systems* 37.5 (2022): 3958-3968.

3 Chapter 5 reproduces the text, with minor changes, of the following published paper: Nicolas Stevens, Yves Smeers and Anthony Papavasiliou. "Indivisibilities in investment and the role of a capacity market." *Journal of Regulatory Economics* (2024), 66:238–272.

4 See also the analysis of "peak-load pricing under indivisibility constraints" by Williamson (1966) who stresses that in a system "with indivisible plant, the fully adjusted long-run static equilibrium can be one in which either positive or negative net profits are realized despite (discontinuously) constant returns to scale. Only accidentally will the enterprise earn zero profits at the welfare maximum."

this problem by means of a capacity investment model with indivisibilities. The approach that is followed is to leverage the tools and the analytical framework developed in chapter 3 and to apply them to the investment problem. The chapter formalizes the issues arising from such a problem and then studies several possible solutions to it, focusing in particular on the favourable role that a so-called "capacity market" *could* play in these settings.

Each of the four chapters is written in such a way that it is self-contained. From time to time, this leads to the repetition of certain concepts and ideas, but hopefully to the benefit of the reader.

2

ELECTRICITY ECONOMICS FUNDAMENTALS

This chapter introduces the main concepts and mathematical models that are employed along this thesis. Since our work is about *electricity markets*, the first two sections revisit the specificities of *electricity* as a commodity and the steps that led to the existence of a *market* for power. Sections 2.3 to 2.6 then outline the main building blocks of an electricity market. Section 2.3 introduces the commodities traded in these markets. Section 2.4 and 2.5 present the consumer model as well as several models for producers that underlie the developments of chapters 3 to 5. Section 2.6 studies the main properties of the allocation and the price of the aforementioned commodities when they are traded in a *competitive* market. As the thesis will study one market failure, namely the non-convexities in the market, these sections also provide the reader with a useful benchmark to keep in mind when reading the subsequent analysis of the thesis. Finally, as electricity markets turn out to be practically organised in a very specific manner, section 2.7 summarizes the overall architecture of power markets. This will also be of importance for the remainder of the thesis, since the pricing rules that will be studied are employed in actual electricity auctions.

2.1 SPECIFICITIES OF ELECTRICITY AS A COMMODITY

Electricity is a peculiar commodity that involves several important and often unique attributes. According to Joskow (2003), "the failure to carefully integrate these attributes into the design of regulatory and market institutions has created market performance problems". The paramount economical and physical features of the electricity sector are the following, classified according to what are the three main economic activities: consumption, production and exchange[1].

DEMAND-SIDE SPECIFICITIES

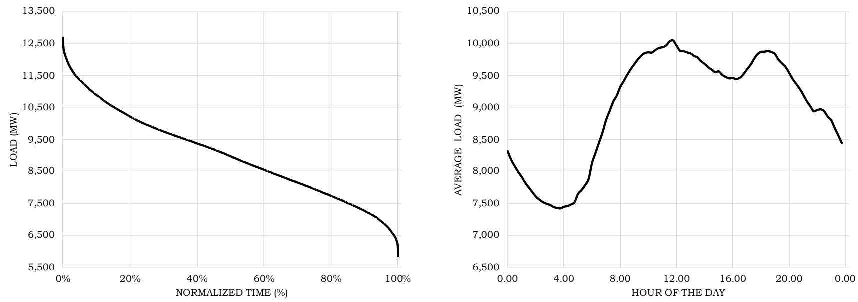
- The demand of electrical energy is highly *inelastic*, especially when it comes close to real-time. This results from the fact that electricity has few substitutes besides temporal substitution. It is also a consequence of electricity demand being, for the most part, disconnected from the market: consumers either are not exposed to the wholesale price of electricity (because retail pricing is flat), or *could not* be exposed to it, as they might not be equipped with the metering technology (hourly meters) that would enable measuring hourly consumption. The inelasticity of electricity consumption is occasionally put forward as one of the most severe flaws of electricity markets (Stoft, 2002). Indeed, increasing demand elasticity (mobilizing demand-side flexibility) would come with multiple benefits, such as a possible reduction in the need to invest in peaking units or the mitigation of market power.
- The demand of electricity *fluctuates significantly* from hour to hour. Figure 2.1 illustrates this fluctuating pattern with the Belgian load in 2023. Hourly load in Belgium can be as high as 12,500MW and as low as almost 5,500MW. As further illustrated in Figure 2.1b, these fluctuations are not only seasonal, but there are also significant variations within a single day—with a typical offpeak at night, a morning peak coming with a steep ramp, and an evening peak.

SUPPLY-SIDE SPECIFICITIES

- The supply merit order also becomes inelastic when approaching the capacity constraint. The merit order curve of the market has a "hockey stick" shape: it tends to become vertical at the end of the merit order, exhibiting an infinite upward leap.[2]
- The production of electrical energy comes with stringent physical constraints, linked to power plant operation, which translate to a supply curve that does not respect the standard assumptions of microeconomics. This feature will be important for our work that is concerned with the non-convexities of the supply curve.

1 For a discussion of these attributes, see in particular Joskow (2003, 2008), Stoft (2002), Borenstein and Bushnell (2000) and Wilson (2002).

2 See (Stoft, 2002, chap. 1.6).



(a) Load duration curve. (b) Average daily profile.

FIGURE 2.1: Belgian load in 2023.

SPECIFICITIES OF THE EXCHANGE OF ELECTRICITY

- Electricity is *exchanged through an electric grid*: unlike most commodities, electricity cannot be exchanged through ships, roads or railways. This electric grid comes with physical constraints that govern the flow of electricity and so the exchange of electricity between locations.

- Electricity *cannot be stored efficiently*. To provide the reader with an order of magnitude, a Tesla Powerwall 2 has a storage capacity of 13.5kWh. Assuming an average wholesale electricity price of 100 €/MWh, 13.5kWh translates into a value of 1.35€, approximately the cost of 1kg of rice. The latter may conveniently be stored in a jar that costs a few euros, while the Tesla Powerwall costs 7,000€.

- *Demand must be met just-in-time* by production. This results from the physical laws governing the electrical grid: demand must be cleared continuously at each location of the network, at any moment. Unlike many other goods, a local shortage of power may not only result in a local increase of electricity prices, but, in the worst case, it may result in cascading outages, meaning the inability to serve the demand, not only locally, but across the entire system.

- *Power cannot be tracked*: consumers can consume power in real time without having an explicit contract to do so.

As we shall particularly study the price of electricity, let us notice that these features translate into a highly volatile price. Table 2.1 presents the statistics of the Belgian day-ahead market price for the years 2016 to 2023. The average price of electricity was 44€/MWh during the years 2016 to 2019, although there have been occurrences of hourly prices as low as

	2016	2017	2018	2019	2020	2021	2022	2023
Hourly min	−5	−41	−32	−500	−115	−70	−100	−120
Hourly max	696	331	499	121	200	620	871	330
Hourly mean	37	45	55	39	32	104	245	97
Hourly perc. 5%	15	22	26	18	9	23	56	6
Hourly perc. 95%	72	85	89	62	56	265	500	170
Daily min	8	11	5	−134	−18	−15	6	−2
Daily max	131	123	185	86	75	433	700	205

TABLE 2.1: Day-ahead electricity prices in Belgium. All figures are in €/MWh. [Data source: ENTSO-E Transparency Platform]

−500€/MWh to as high as 696€/MWh. Moreover, in 2019 there was an entire day with an average price of −134€/MWh, while in 2018 there was a day with an average price of 185€/MWh. In 2022, during the gas crisis, the price culminated at 871€/MWh, with one instance of an average daily price of 700€/MWh!

2.2 MARKETS FOR POWER: THE DEREGULATION OF POWER SYSTEMS

RESTRUCTURING OF THE ELECTRICITY SECTOR. The supply chain of electricity is made of four main components[3] (i) power generation, (ii) transmission and (iii) distribution grids, and (iv) retail (Joskow, 2008). These components used to be organised all together in a vertically integrated monopoly, called "utility". Utilities were either privately owned and subject to the regulation by the State or, alternatively, publicly owned. This is still the case in many regions of the world. However, since the 1990s, liberalization policies took place in most western countries, moving from *centralized* operations to *decentralized* market mechanisms. Alternatively phrased, economic activities that used to be coordinated by a *firm*, where the usage of resources essentially resulted from administrative decisions, moved to coordination through *market transactions* which relies on the price signal to allocate resources (Coase, 1937). This institutional change led in particular to:[4]

- A "vertical" separation of the *competitive* segments—generation and retail—from the *natural monopoly* segments—transmission and dis-

3 The topic of security of supply for fuels is out of the scope of the chapter.
4 See the analysis of Joskow (2003, 2008) and Borenstein and Bushnell (2015).

tribution. Electricity networks indeed exhibit a "natural monopoly" property: they are characterized by decreasing average cost, as well as by ubiquitous externalities. For instance, due to the laws of physics, the transport of energy on one line has a direct implication on what is transported on another line (Borenstein and Bushnell, 2000), cf. section 2.6.2.

More specifically, this "unbundling" involves three main restructuring policies: the restructuring of the electrical grid (the independent oversight of the network), the restructuring of generation ownership (the divestiture of existing State-owned assets and the free entry of unregulated plants) and the restructuring of retail. Although these three reforms are in principle intertwined, they have not always come together in practice.[5]

- A "horizontal" integration of networks, or a "coupling" between regions. In Europe, this concretely translates into (roughly) one TSO per country that monitors the entire transmission grid and a coupling of these national networks through European-scale markets (an important piece being the day-ahead market also called "Price Coupling of Regions").
- The creation of electricity wholesale markets.
- The creation of independent regulatory agencies and regulatory mechanisms, an "underappreciated component of the successful reforms" according to Joskow (2008).

GOALS OF DEREGULATION. Why deregulate?

> "One might ask why bother with the difficult process of creating wholesale electricity markets with these attributes if we are simply reproducing the central planning results for generator scheduling and dispatch? The answer is that the central planning models for vertically integrated utilities are 'idealized' models that do not take into account the incentives faced by the regulated vertically integrated monopoly and how these incentives affect behaviour. It is generally thought that regulated monopolies have poor incentives to control operating and construction costs, to maintain generator availability at optimal levels, to retire generators when the expected present value of their costs exceeded the expected present value of continuing operations, to overinvest in new generating capacity, to fail

5 See Borenstein and Bushnell (2015) for a discussion and for an analysis of the US.

> aggressively to seek out innovations, and other inefficiencies. In short, the real world regulated monopoly does not perform as the idealized model implies." (Joskow, 2019)

In other words, there were inefficiencies in the "old regime" that the market was meant to solve. In particular, the efficiency gains argued in favour of deregulation are of two sorts: *short-term* efficiency and *long-term* efficiency (Cramton, 2017). Promoting efficiency in short-term operations means providing incentives to producers to control their production cost. Long-term efficiency means inducing efficient investments. This has often been argued to be the main benefit from deregulation: "Most efficiency gains from restructuring will be long-term resulting from better investment decisions" (Oren, 2000). Aligned with these two objectives, deregulation also aimed at promoting innovation and making sure that the prices that are observed by consumers reflect the true cost of production, so as to accurately signal scarcity and induce an efficient usage of factors by consumers.

The theoretical arguments supporting the alleged efficiency of competitive markets for power are outlined in section 2.6. However, "*deregulation is not equivalent to perfect competition*" (Stoft, 2002): merely deregulating does not necessarily lead to the economic efficiency of a competitive market. In practice, markets are imperfect. For example, there may be exertion of market power. This might especially be a concern as several features of electricity might exacerbate market power (Joskow, 2008; Borenstein and Bushnell, 2000): inelasticity of demand ; the geographical limitation of competition due to a tight network ; or the concentration of generation capacity in the hands of a few firms (see, for instance, the concentration of plant ownership in Belgium in Figure 2.2). Deregulation may also create new challenges: if *over*-investment was a major concern in vertically integrated monopolies with regulatory-driven investments—a concern that the market was meant to solve—, *inadequacy* (or under-investment), has become the new challenge with market-driven investments (Borenstein and Bushnell, 2000). Therefore, a more nuanced view of deregulation might be phrased as follows: "the move to liberalizing the electricity sector in this way was effectively a bet that the costs of any residual imperfections in competitive wholesale markets are smaller than the costs of imperfections associated with the behaviour of vertically integrated regulated monopolies." (Joskow, 2019)

PERFORMANCE OF DEREGULATION. After more than twenty years, what is the empirical evidence of the successes of deregulation policies? Some early assessments were not necessarily optimistic. According to Borenstein and Bushnell (2000), "Probably the two most salient lessons are that the

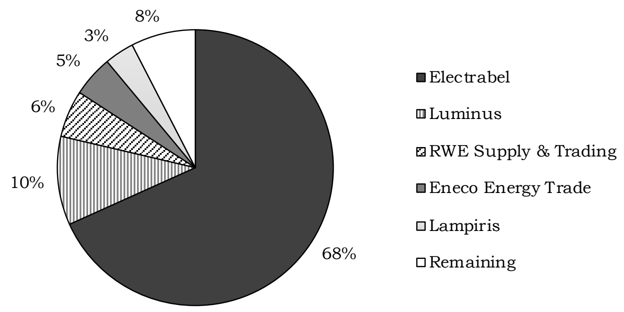


FIGURE 2.2: Ownership of the generation mix in Belgium. [Data source: Elia website]

short-run benefits are likely to be small or nonexistent, and the long-run benefits, while compellingly supported in theory, may be very difficult to document in practice." However, more recent evidence suggests certain significant efficiency improvements due to deregulation. Schmalensee (2021) summarises the main data. Empirical evidence in the US highlights that there were significant operating cost improvements in deregulated areas, in particular for nuclear plants (Davis and Wolfram, 2012), but also for a broader set of technologies (Fabrizio et al., 2007)[6]. Other evidence highlights the significant efficiency gains resulting from increasing cross-border trades and nodal pricing, permitted by the creation of organised wholesale markets[7]. These savings in operating costs may however be counterbalanced by the departure of price from marginal production cost, due to the exercise of market power. Borenstein and Bushnell (2015) also highlights that the exposure to gas prices is an unintended consequence of deregulation. The price, in the vertically integrated arrangement, is the *average cost* of electricity which is typically not highly correlated with gas price as gas-fired power plants are not the dominant producing technology. Instead, in a liberalized market, the price equals *marginal cost*, which is highly correlated with the gas price—as Europe dramatically experienced during the gas crisis in 2022. The overall appraisal of Schmalensee (2021) is that the benefits of deregulation are positive, although not dramatic.

6 Davis and Wolfram (2012) find a 10% increase in operating performances, mainly driven by a reduction of outage duration. Fabrizio et al. (2007) find an improvement of 6% (resp.12%) in labor use (resp. non-fuel expenses) of restructured plants with respect to government-owned plants.

7 This is also emphasized by Borenstein and Bushnell (2015): "The creation and expansion of the RTO/ISO model may be the single most unambiguous success of the restructuring era in the United States. [...] The evidence suggests that the lack of coordination across utility control areas impeded Pareto-improving trades worth billions of dollars".

2.3 COMMODITIES

The *ultimate* goal, or the main service, delivered by the entire electricity sector is *the provision of electrical energy in real time to the end-consumer*. Achieving this goal, however, requires the exchange of several physical commodities, as well as numerous financial products through which the physical commodities are traded on different time horizons. These practicalities and institutional arrangements will be discussed in section 2.7 and will be ignored in the present section which focuses on the actual physical commodities. The three main physical commodities exchanged in the electricity sector are: (i) electrical energy, (ii) transmission capacity and (iii) various ancillary services. These three commodities should further be thought of as also being indexed by time and location: energy at time t and in location i, etc.

Transmission capacity stand for the right to use the network in order to transport electrical energy from producers to consumers: to buy energy in A from B, one should buy both the energy produced in B as well as the right to transport it on the grid from B to A. The various ancillary services ensure power quality and contribute to the reliability of the system. Their need is driven by the features of electricity developed in section 2.1: because electricity demand is highly fluctuating, because energy cannot be stored efficiently and because, despite this, consumption should equal production at each and every instant, together with the stringent physical constraints of the grid, this justifies the existence of these services. They include various products such as frequency control, voltage control or black-start services.[8] Abstracting from some (important) subtleties, it is convenient for the purpose of the present chapter to limit our analysis of these services to an aggregate "reserve", or a "real-time stock of power":

8 (i) *Frequency control* and, more generally, *balancing service* is the service of balancing the system, which translates into the continuous control of frequency at the customer endpoint. It encompasses the variety of reserve products such as Frequency Curtailment reserve (FCR), Automatic/Manual Frequency Restoration Reserve (aFRR and mFRR), etc. (ii) *Voltage stability* is managed through the provision of reactive power which is typically an out-of-market arrangement: since reactive power does not transport easily, a deregulated market would be exposed to market power (Stoft, 2002, p. 21). (iii) *Black-start capability* is the service contracted by the system operator in order to reactivate the system after a black-out, since most of the units cannot start without a functioning system.
On top of these three services, for which the system operator is the buyer, Stoft (2002), chap. 3–4, also includes the following services for which the system operator is the sole provider. (iv) *Trading enforcement* which implies the metering of the injections and withdrawals, as well as monitoring of the power flows. Since power cannot be tracked, an *independent* party should measure the injections and withdrawals of power for contractual purposes. (v) *Economic dispatch*. This service includes providing dispatch and (in day-ahead) commitment instructions to the suppliers.

Commodity	Type of good	Supply-side	Demand-side
Electrical energy	private good	deregulated	deregulated
Transmission capa.	private good	regulated (*SO is monopoly supplier*)	deregulated
Ancillary services	public good	deregulated	regulated (*SO is monopsony buyer*)

TABLE 2.2: The three main commodities exchanged in the electricity sector.

some "spare capacity" $r_{g,t}$ left available by the supplier g at time t in order for the system operator to have the flexibility to cope with contingencies.

From an economic viewpoint, these three commodities hold very different properties. Electrical energy is a classic *private good*. Conversely, ancillary services are typical examples of *public goods*: they are *non-rival* (the consumption by one agent of the good—the "reliability" of the system or "quality" of power, for instance—would not prevent another agent from consuming it) and *non-exclusive* (an agent connected to the grid may hardly be prevented to consume this good or be charged for it). They are exposed to the classical free-riding problem which would lead to an undersupply of these goods, were they not regulated. Transmission—the "roads" of electricity—is a natural monopoly because of economies of scale. The monopolist transmission operator then provides the market participants with well-defined and enforceable rights of using the grid, making the transmission capacity a private good (both rivalrous and exclusive).

These differences lead to heterogeneous market arrangements, which are either deregulated or partially deregulated. The market for energy has both a supply-side and a demand-side which are deregulated to some extent[9]. The market for transmission capacities has a regulated supply-side—the system operator is the *monopoly* seller of transmission capacity—while the demand-side is deregulated. Finally, the market for reserve has a deregulated supply-side—reserve is supplied by private power plants—while it has a regulated demand-side: the system operator is the *monopsony* buyer of these services. Table 2.2 summarizes the main features.

9 By "deregulated" we do not mean a complete *laisser-faire* but simply that the demand or supply side rely on individual decisions, as opposed to a *regulated* regime which is driven by administrative decisions.

2.4 CONSUMER THEORY

Throughout our work, we adopt a *partial equilibrium* analysis (Mas-Colell et al., 1995, chap. 10). That is, we ignore cross-market effects. For the suppliers, we assume that variable production costs and investment costs are given: fuel prices—gas, oil, uranium, etc.— as well as input of investments costs—steel and concrete prices, labour, etc.—are assumed to be unaffected by electricity prices. For consumers, we assume that income (or wealth) is given.

Let $((d_l)_{l \in \mathcal{L}}, z)$ be a consumption bundle. $(d_l)_{l \in \mathcal{L}}$ is the set of commodities we are interested in: electricity consumption in $l \in \mathcal{L}$ ($\mathcal{L}$ should be viewed as the Cartesian product of the sets of locations and time periods). z is the *numeraire*—the "Hicksian composite commodity": a composite commodity standing for all the other commodities in the economy that we do not study. The numeraire price is normalized to 1: it is the reference, or the "money value", towards which the value of commodities d_l is measured. Each consumer $j \in \mathcal{J}$ is assumed to hold a preference relation $\succsim_j$, defined on the commodity space, which is represented by a utility function $u_j : \mathbb{R}^{L+1} \to \mathbb{R}$. We shall assume that the utility function takes a *quasi-linear* form: $u_j(d_j, z_j) = v_j(d_j) + z_j$. Since the price of the numeraire z_j is normalized to 1, $v_j(d_j)$ may be viewed as the monetary value that consumer j assigns to bundle d_j, or his *willingness-to-pay* for bundle d_j (recall that d_j is a vector of L goods). The consumer's decision problem is then the following:

$$\max_{d_j, z_j \geq 0} v_j(d_j) + z_j$$

$$\pi^T d_j + z_j \leq w_j$$

Here, w_j is the wealth of consumer j and π is the vector of electricity prices. Assuming locally non-satiated preferences, the wealth constraint is tight and $z_j = w_j - \pi^T d_j$. The consumer decision problem then simplifies to:

$$u_j^* = w_j + \max_{d_j \geq 0} \{v_j(d_j) - \pi^T d_j\} \tag{2.1}$$

This is the consumer surplus maximization problem. As we may observe, following our partial equilibrium assumptions, the problem is independent of wealth distribution—there is no *wealth effect*: a change in the wealth of consumer j does not change the consumption decision d_j. This results from the assumption of quasi-linearity of the utility function. In this model, variations of consumer *welfare*, for example following a change in market price, are measured in variations of consumer *surplus* in the electricity market (equation (2.1)).

$d_j(\pi)$, the solution of problem (2.1), is the *demand correspondence* of j and $d(\pi) = \sum_{j\in\mathcal{J}} d_j(\pi)$ is the *aggregated demand correspondence*. In most of this work, we assume that $v(d) = \sum_{j\in\mathcal{J}} v_j(d_j)$ takes the following linear form: d is valued at $VOLL$ (the Value of Lost Load) from 0 to D (the total load), and valued at 0 for $d > D$. That is, the aggregate decision problem is:

$$\max_{d} \sum_{l\in\mathcal{L}} d_l VOLL - \pi^T d \tag{2.2a}$$

$$0 \leq d_l \leq D_l \qquad \forall l \in \mathcal{L} \tag{2.2b}$$

The $VOLL$, a concept broadly used in the electric industry, should be viewed as the *system-wide willingness-to-pay* for electricity (or the system-wide willingness-to-pay to avoid a power outage). This is obviously a simplification of reality, in which the willingness-to-pay for electricity depends on *end-user* (residential, industrial, a person in particular, etc.), *use-case* (consumers typically do not value electricity directly, but the end-usage whose electricity is an input: medical appliances, lightening, etc.) and *consumption context* (outage duration, local or global outage, warning message, etc.).[10]

All-in-all, model (2.2) is a stylized view of consumers. It is however a commonly adopted model in electricity markets, justified by the inelastic nature of electricity consumption. As this thesis will focus on the non-convexities that are present on the supply curve, it is also convenient to adopt such a stylized model for consumption. We shall occasionally adopt, during the thesis, an even simpler model in which electricity demand is fully inelastic.

2.5 PRODUCER THEORY

Let us consider a set of $\mathcal{G}$ suppliers. Each supplier is located at a certain node of the grid and produces an amount $q_{g,t}$ of energy at time t. Its production cost for a certain production plan $q_{g,t}$ over the time horizon $\mathcal{T}$ under study is c_g. The production set is denoted as $(c_g, q_{g,t}) \in \mathcal{X}_g$. Each supplier is assumed to maximize its profit given an electricity price π_t, that is:

$$\max_{(c,q)_g \in \mathcal{X}_g} \sum_{t\in\mathcal{T}} \pi_t q_{g,t} - c_g \tag{2.3}$$

10 Furthermore, estimating the $VOLL$ turns out to be complex. Many approaches exist such as techo-engineering proxy, contingent valuation studies or revealed preference approaches. It is worth stressing that differences of several orders of magnitude have been reported across studies. The estimation of the $VOLL$ is further complicated by the public good nature of electric reliability. See Gorman (2022) for a detailed discussion.

In most of this thesis, we remain general and simply denote the production set by $\mathcal{X}_g$, occasionally specifying whether $\mathcal{X}_g$ is convex (as problem (2.4) below) or non-convex (as models (2.5) or (2.7) below). This section aims at giving a sense to the reader of what is within this production set $\mathcal{X}_g$.

CONVEX PRODUCTION MODEL (ECONOMIC DISPATCH). The most elementary convex electricity production model is the following. The output $q_{g,t}$ produced by a supplier of electricity g at time t is limited by the installed capacity of the power plant Q_g^{max}. It also comes with a production cost function which is assumed to be linear: $C_g(q_{g,t}) = MC_g q_{g,t}$. The model of the production set $\mathcal{X}_g$ then corresponds to equations (2.4).

$$c_g = \sum_{t \in \mathcal{T}} MC_g q_{g,t} \tag{2.4a}$$

$$0 \leq q_{g,t} \leq Q_g^{max} \qquad \forall t \in \mathcal{T} \tag{2.4b}$$

UNIT COMMITMENT MODEL. Model (2.4) does not account for some important features and constraints that are present in production processes. In reality, power plants have to be operated between bounds, incur some fixed costs when they are started up, are limited in their ramping capability, or cannot be turned on and off too frequently. Therefore, the following more comprehensive model called the "unit commitment model" is often adopted to represent production more accurately.

$$c_g = \sum_{t \in \mathcal{T}} \left(MC_g q_{g,t} + NC_g u_{g,t} + SC_g v_{g,t} \right) \tag{2.5a}$$

$$Q_g^{min} u_{g,t} \leq q_{g,t} \leq Q_g^{max} u_{g,t} \qquad \forall t \in \mathcal{T} \tag{2.5b}$$

$$u_{g,t} = u_{g,t-1} + v_{g,t} - z_{g,t} \qquad \forall t \in \mathcal{T},\ t > 1 \tag{2.5c}$$

$$q_{g,t} \leq q_{g,t-1} + R_g^+ - v_{g,t}(R_g^+ - R_{SU}^+) \qquad \forall t \in \mathcal{T},\ t > 1 \tag{2.5d}$$

$$q_{g,t} \geq q_{g,t-1} - R_g^- + z_{g,t}(R_g^- - R_{SD}^-) \qquad \forall t \in \mathcal{T},\ t > 1 \tag{2.5e}$$

$$\sum_{i=t-UT_g+1}^{t} v_{g,i} \leq u_{g,t} \qquad \forall t = UT_g, ..., T \tag{2.5f}$$

$$\sum_{i=t-DT_g+1}^{t} z_{g,i} \leq 1 - u_{g,t} \qquad \forall t = DT_g, ..., T \tag{2.5g}$$

$$u_{g,t}, v_{g,t}, z_{g,t} \in \{0, 1\} \qquad \forall t \in \mathcal{T} \tag{2.5h}$$

Three new variables are introduced: $u_{g,t}$, which is the commitment variable indicating whether the supplier g is online or offline at period t; $v_{g,t}$, which is the start-up variable indicating whether g is switched on at time t and $z_{g,t}$, the shut-down variable, indicating whether g is switched off at time

t. Constraint (2.5c) links these three variables logically. The main novelty with respect to model (2.4) is the presence of non-convexities in the model. These non-convexities materialise in the cost function (2.5a) through some fixed production costs: a no-load cost NC_g and a start-up cost[11] SC_g. They also appear in the operating limits of the plants, in constraint (2.5b): the plant either produces 0 or is contained between a minimum Q_g^{min} and a maximum Q_g^{max}. The last two sets of constraints include the so-called minimum up and down time constraints (2.5f)–(2.5g) and the ramping constraints (2.5d)–(2.5e).

EUROPEAN MARKET MODEL. As it will be occasionally mentioned later in this thesis, it is useful to introduce here a third model of supply that we shall refer to as the "European market model". A preliminary remark is needed to avoid any confusion. It is important to distinguish two things when analysing an electricity market: the conceptual models that can be used to describe the functioning of the market and the models actually employed in the concrete organization of the marketplace. Electricity markets are indeed very peculiar for the highly centralized way they are organised as closed-gate auctions (this will be later explained in section 2.7). These auctions typically involve solving highly complex optimization models. The unit commitment model accurately describes the functioning of the supply-side of an electricity market. This holds true in the United States, but also in Europe. The unit commitment model also turns out to be a widely employed model for the actual electricity auctions held in the US. This is an important difference with Europe, in which the day-ahead auction employs an alternative model which relies on totally different bidding products. It is the latter that we briefly describe here.

Ignoring the so-called "complex orders" and the specific Italian orders called the "PUN" (for "Prezzo Unico Nazionale")[12], the bid constraints in the European day-ahead market can be split in two categories (NEMO Committee, 2020b): (i) *convex* orders that include hourly orders (stepwise

11 Start-up costs typically have two main sources: first, an extra fuel consumption that is incurred to warm up the power plant and to switch it on; and second, the cost of an overhaul to be performed whenever the power plant has been switched on a certain number of times.

12 We neglect these bidding products as they are planned to be discontinued (MCSC, 2023). The overall idea of the PUN is as follows. Italy is represented in the European day-ahead auction with multiple zones. Nonetheless, Italian stakeholders have expressed the following—highly disputable, in my opinion—request: it is deemed acceptable for Italian suppliers to face different prices across Italy, but the demand should face the same price everywhere in Italy. Thus, the Italian demand orders—the so-called PUN orders—shall be cleared at the "PUN price" (in Italian, the "Prezzo Unico Nazionale") instead of the bidding zone price. The PUN price should be such that the collected payments from the demand cover the expenses to Italian suppliers that are subject to the regular bidding zone prices. Mathematically, if d_z denotes the cleared demand in Italian zone z, q_z denotes the cleared production in zone z, π_z

and interpolated) and price-taking orders; and (ii) *non-convex* orders which include families of block orders. Stepwise orders are similar to model (2.4): for each time period, these orders specify a maximum amount of production $Q_{g,t}$ at a given price $P_{g,t}$ which may be continuously accepted, so $(c_g, q_{g,t}) \in \mathcal{X}_g \equiv \{(c,q) \mid c_g = \sum_t P_{g,t}q_{g,t},\ q_{g,t} = x_{g,t}Q_{g,t},\ 0 \leq x_{g,t} \leq 1\}$. Variable $x_{g,t}$ stands for the continuous acceptance ratio of the bid quantity $Q_{g,t}$. Interpolated orders are similar in principle, except that the total cost is quadratic instead of linear. The block order constraints $\mathcal{X}_g$ of a block g can be described as follows:

$$
\begin{aligned}
c_g &= \sum_{t \in \mathcal{T}} P_g q_{g,t} && \text{[block order price]} && (2.7a)\\
q_{g,t} &= Q_{g,t} x_g \quad \forall t \in \mathcal{T} && \text{[production of the block]} && (2.7b)\\
0 \leq x_g &\leq 1 && \text{[block continuous acceptance]} && (2.7c)\\
u_g &\in \{0,1\} && \text{[block discrete acceptance]} && (2.7d)\\
x_g &\geq u_g R_g && \text{[block “fill” (min ratio) const.]} && (2.7e)\\
x_g &\leq u_g && \text{[block “kill” const.]} && (2.7f)\\
x_g &\geq x_{g^2} \quad \forall g^2 \in Child(g) && \text{[parent-child const.]} && (2.7g)\\
\sum_{g' \in Excl} x_{g'} &\leq 1 \quad \text{if } g \in Excl && \text{[exclusive group const.]} && (2.7h)
\end{aligned}
$$

Compared to the stepwise and interpolated orders, the block constraints introduce an additional variable u_g which stands for the *binary* acceptance ratio of the block. Let us notice that constraints (2.7g)–(2.7h) link block g with other blocks through “parent-child” (or linked block) relationships (eq. (2.7g)) and exclusive group relationships[13] (eq. (2.7h)). Let us also notice that the acceptance of a block is not indexed by time: the profile $Q_{g,t}$ (a parameter of the bid) spans over multiple periods, and the quantity

denotes the price in zone z and π^{PUN} denotes the (unique) PUN price, then the following relationship should hold:

$$\pi_t^{PUN} \sum_{z \in Italy} d_{z,t} = \sum_{z \in Italy} \pi_{z,t} q_{z,t} \qquad \forall t \in \mathcal{T} \qquad (2.6)$$

Mathematically, expression (2.6) is a *primal-dual constraint*: it involves both *primal* variables q and d as well as *dual* variables $\pi_{z,t}$ and π_t^{PUN}. From an algorithmic standpoint, Euphemia, the European market clearing algorithm, first solves the primal problem and finds a candidate allocation q^* and d^*. Then, it attempts to find prices $\pi_{z,t}$ and π_t^{PUN} that satisfy constraint (2.6). If no price can be found, the algorithm adds a cut in the primal model and repeats the process. This primal-dual constraint turns out to be particularly challenging from a computational standpoint.

13 Note that the exclusive group constraint is imposed on the continuous variable x_g, which means that, if the blocks have an acceptance ratio R_g smaller than 1, there could possibly be multiple blocks accepted.

offered may be different between periods, but the acceptance rate x_g is equal for all the time periods.

2.6 COMPETITIVE ANALYSIS

This section outlines the fundamental notions of a competitive analysis applied to electricity markets.[14] This analysis is important to understand why it *could* be true that using a market to coordinate the economic activities of the electricity sector would lead to an allocation of resources that may be regarded as "optimal" in a well-defined sense. Also, as Arrow and Hahn emphasize in the introduction of their book "In attempting to answer the question 'could it be true', we learn a good deal about why it might not be true." (Arrow and Hahn, 1971, p. vii) In other words, the competitive analysis offers a useful benchmark from which inefficiencies due to market failures—such as *non-convexities*—can be traced and eventually corrected by regulatory intervention.

The subsequent analysis in chapter 3 could later fruitfully be read in contrast with the analysis of this section. While the present section will consider the consumer model (2.2) and the *convex* production model (2.4), the analysis of chapter 3 will study what happens when, instead, *non-convex* production models such as (2.5) or (2.7) are adopted.

2.6.1 *Competitive equilibrium*

The main value judgement we will employ along this thesis to evaluate policies or outcomes of markets is the notion of *economic efficiency*. We are interested to reach an allocation of resources that is economically efficient. In certain cases, this will mean minimization of costs. In others, it will mean maximization of the market surplus. But before examining such a problem below (problem (2.9)), it is worth briefly recalling its connection with the fundamental concept of Pareto optimality.

---

14 Throughout the thesis, our modelling methodology will mostly rely on competitive analysis. That is, we will neglect strategic behaviors. There are two main motivations for this choice. First, as we shall see, the issue of pricing non-convexities turns out to be challenging, even in competitive settings. Thus, our work attempts to address this problem first in competitive settings, while future works could extend it to include strategic behavior. Secondly, this choice enables us to use tools from linear programming and mixed-integer programming, which allow to model the complexity of electricity auctions in fine details (while relying on e.g. game theory would likely require many simplifying assumptions regarding the bidding complexity, the treatment of the network, etc.).

Definition 2.1 (Pareto Optimality)**.** *A feasible allocation* (d, z, q) *is Pareto optimal if there is no other feasible allocation* (d', z', q') *such that* $u_j(d'_j, z'_j) \geq u_j(d_j, z_j)$ $\forall j \in \mathcal{J}$ *and* $\exists j : u_j(d'_j, z'_j) > u_j(d_j, z_j)$.

Pareto optimality formalizes the notion of an efficient usage of resources: in a Pareto optimal allocation, it is impossible to use resources in a way that makes someone better off without making someone else worse off. In our partial equilibrium setting, under the assumption of quasi-linear utility functions, if an allocation (d^*_j, q^*) is Pareto efficient then other Pareto-efficient allocations can be obtained by lump-sum redistribution of wealth between agents. This means that the utility possibility set is defined as

$$\{(u_1 ... u_J) : \sum_{j \in \mathcal{J}} u_j \leq \sum_{j \in \mathcal{J}} v_j(d_j) + \underbrace{\sum_{j \in \mathcal{J}} w_j - \pi^T \sum_{j \in \mathcal{J}} d_j}_{\text{wealth left}} \} \tag{2.8}$$

Since the short-term profit of producers may be non-negative, we need to specify where these profits go. We assume a classic private ownership economy in which the wealth of each individual consumer is $w_j = \omega_j + \sum_{g \in \mathcal{G}} \theta_{jg}(\pi^T q_g - c_g)$ (an exogenous endowment ω_j of numeraire plus a share of profit θ_{jg} of supplier g, with $\sum_{j \in \mathcal{J}} \theta_{jg} = 1$ $\forall g$). With these assumptions, the "wealth left" of expression (2.8) simplifies as follows:

$$\begin{aligned} \sum_{j \in \mathcal{J}} w_j - \pi^T \sum_{j \in \mathcal{J}} d_j &= \sum_{j \in \mathcal{J}} \omega_j + \sum_{j \in \mathcal{J}} \sum_{g \in \mathcal{G}} \theta_{jg}(\pi^T q_g - c_g) - \pi^T \sum_{j \in \mathcal{J}} d_j \\ &= \sum_{j \in \mathcal{J}} \omega_j - \sum_{g \in \mathcal{G}} c_g \end{aligned}$$

Then, simplifying expression (2.8), the allocation on the Pareto frontier can be obtained from the following problem:

$$\max_{q,d \geq 0} \sum_{t \in \mathcal{T}} d_t VOLL - \sum_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}} MC_g q_{g,t} \tag{2.9a}$$

$$(\pi_t) \;\; d_t \leq \sum_{g \in \mathcal{G}} q_{g,t} \qquad \forall t \in \mathcal{T} \tag{2.9b}$$

$$(\mu_{g,t}) \;\; q_{g,t} \leq Q_g^{max} \qquad \forall g \in \mathcal{G}, t \in \mathcal{T} \tag{2.9c}$$

$$(\nu_t) \;\; d_t \leq D_t \qquad \forall t \in \mathcal{T} \tag{2.9d}$$

The objective function (2.9a) corresponds to the so-called *Marshallian aggregate surplus* of the electricity market under study. Problem (2.9) aims at finding a production plan q and a consumption plan d such that these plans maximize the aggregate surplus (2.9a), while satisfying production constraints (2.9c) and consumption constraints (2.9d), and such that the market clears (equation (2.9b)). Let us notice that we assume *free disposal*. The optimality conditions of problem (2.9) are:

$$0 \leq q_{g,t} \quad \perp \quad MC_g - \pi_t + \mu_{g,t} \geq 0 \qquad \forall g \in \mathcal{G}, t \in \mathcal{T} \tag{2.10a}$$

$$0 \leq d_t \quad \perp \quad \pi_t - VOLL + \nu_t \geq 0 \qquad \forall t \in \mathcal{T} \tag{2.10b}$$

$$0 \leq \sum_{g \in \mathcal{G}} q_{g,t} - d_t \quad \perp \quad \pi_t \geq 0 \qquad \forall t \in \mathcal{T} \tag{2.10c}$$

$$0 \leq Q_g^{max} - q_{g,t} \quad \perp \quad \mu_{g,t} \geq 0 \qquad \forall g \in \mathcal{G}, t \in \mathcal{T} \tag{2.10d}$$

$$0 \leq D_t - d_t \quad \perp \quad \nu_t \geq 0 \qquad \forall t \in \mathcal{T} \tag{2.10e}$$

There are two paramount observations that can be made from equations (2.10). Firstly, the reader may check that the optimality conditions (2.10) are equivalent to the optimality conditions of problems (2.4) and (2.2) together with the market-clearing condition (2.9b). This *remarkable* fact formalizes the notion that a decentralized market economy in perfect competition reproduces the outcome of a centralized problem maximizing the total surplus of the agents in the market. In other words, the outcome of the market that coordinates the agents by means of a price signal achieves *allocative efficiency*, or *Pareto efficiency*. This is the *first theorem of welfare economics* in a partial equilibrium setting.

Secondly, from the optimality conditions stated above, in order to achieve this efficient allocation of resources, the marginal supplier sets the uniform price of electricity. Indeed, let g^m be the supplier such that, at time t, $0 < q_{g^m,t} < Q_{g^m}^{max}$. Then $\pi_t = MC_{g^m}$, which means *marginal pricing* prevails in the market. Thus an efficient allocation of resources is achieved by a market economy in which resources are priced at their marginal cost of production. This concept of *marginal pricing* will be important for this thesis as non-convexities will challenge its appropriateness in more general settings.

2.6.2 *Competitive equilibrium with a network*

CONVEX NETWORK MODEL. An important feature of the exchange of electricity is that it goes through an electrical grid which has stringent physical constraints (cf. section 2.1). Each consumer and producer is located on a certain node $i \in \mathcal{N}$ of the network. There is a flow f_k on line k connecting nodes i and j. These flows are restricted by network constraints: $f \in \mathcal{F}$. Along this thesis, we will remain general and refer to $\mathcal{F}$ as an abstract set of constraints on the network, that we will however assume to be convex. The present section aims at giving a sense to the reader of what lies in the set $\mathcal{F}$.

The physical laws that govern the flows on an electrical network can be expressed as the following so-called "AC power flow equations" (Taylor, 2015). These equations link the voltages at nodes i (v_i) and j (v_j) with the active (f_k) and reactive (f_k^{react}) power flows on line k connecting i and j (for simplicity, we omit the time index t):

$$f_k = g_k v_i^2 - v_i v_j (g_k \cos(\theta_i - \theta_j) - b_k \sin(\theta_i - \theta_j)) \tag{2.11a}$$

$$f_k^{react} = b_k v_i^2 - v_i v_j (g_k \sin(\theta_i - \theta_j) + b_k \cos(\theta_i - \theta_j)) \tag{2.11b}$$

$$\forall k \in \mathcal{K},\ i = or(k),\ j = dest(k)$$

with g_k and b_k being respectively the *conductance* and the *susceptance*, i.e. the real and imaginary part of the *admittance*[15] defined as $y_k = g_k - \iota b_k$ where $\iota^2 = -1$; $or(k)$ and $dest(k)$ denote the origin and destination nodes of line k. The active and reactive power balance (or market clearing) constraints at each node i are:

$$\sum_{g \in \mathcal{G}_i} q_g - d_i = \sum_{k \in from(i)} f_k - \sum_{k \in to(i)} f_k \qquad \forall i \in \mathcal{N} \tag{2.12}$$

$$\sum_{g \in \mathcal{G}_i} q_g^{react} - d_i^{react} = \sum_{k \in from(i)} f_k^{react} - \sum_{k \in to(i)} f_k^{react} \qquad \forall i \in \mathcal{N} \tag{2.13}$$

with $\mathcal{G}_i$ being the set of suppliers in i, q_g^{react} and d_i^{react} denoting the supply and demand of reactive power in i; $from(i)$ and $to(i)$ denoting respectively the set of lines that flow from i and to i. Finally, the box constraints on the voltage and on the apparent power flow are:

$$\underline{v_i} \leq v_i \leq \overline{v_i} \tag{2.14}$$

$$(f_k)^2 + (f_k^{react})^2 \leq \overline{f_k}^2 \tag{2.15}$$

Constraints (2.11) introduce another source of non-convexities in the electricity market. However, we shall neglect them in this thesis and adopt instead a *linear* model of the network. This has three main motivations. Firstly, we are mainly interested in the non-convexities on the supply-side and how the electricity market can treat them. Therefore, it is convenient to limit the complexity of the network model. Secondly, electricity auctions, as they are implemented in the US and in Europe, typically assume linear power flow equations. Thirdly, the linear power flow model is actually a reasonable approximation for the high-voltage grid.

The *linear* power flow model can be constructed as follows. Let us assume that (i) voltage magnitudes are close to 1 p.u., that (ii) $g_k << b_k$, and that (iii)

15 The admittance can be also be obtained from the line *impedance* $z_k = r_k + \iota x_k$, where the real part is the *resistance* and the imaginary part the *reactance*, as $y_k = 1/z_k$.

the voltage angle differences between the nodes are small, which implies $\sin(\theta_i - \theta_j) \sim (\theta_i - \theta_j)$. As a consequence of these assumptions, the reactive power as well as the voltage terms may be neglected and constraints (2.11)–(2.15) may be rewritten as the following linear constraints[16]

$$f_k = b_k(\theta_i - \theta_j) \qquad \forall k \in \mathcal{K},\ i = or(k),\ j = dest(k) \tag{2.16}$$

$$-\overline{f_k} \leq f_k \leq \overline{f_k} \qquad \forall k \in \mathcal{K} \tag{2.17}$$

Our network model is then:

$$f \in \mathcal{F} \equiv \{f | \exists (f, \theta) \text{ satisfying constraints (2.16)–(2.17)}\}.$$

ECONOMIC IMPLICATIONS. The analysis of the locational pricing of electricity was pioneered by Bohn et al. (1984). The objective of this short section is not to provide a comprehensive analysis of locational pricing, but to highlight some of the fundamental properties that are good to bear in mind for the next chapters (see Papavasiliou (2024) and Taylor (2015) for a more extensive coverage of the topic). Let us assume a basic network with two nodes $\mathcal{N} = \{A, B\}$ connected by a line, where the flow f is defined as positive when energy moves from A to B. The linear network model (equations (2.16)–(2.17)) then simplifies to the capacity constraints on the line for this stylized two-nodes network[17]. For simplicity, we omit the time index t again, although the analysis is straightforward to generalize to multi-time market.

$$\max_{q,d \geq 0, f} \sum_{i \in \mathcal{N}} d_i VOLL - \sum_{g \in \mathcal{G}} MC_g q_g \tag{2.18a}$$

$$(\pi^A) \quad \sum_{g \in \mathcal{G}^A} q_g - d_A = f \tag{2.18b}$$

$$(\pi^B) \quad \sum_{g \in \mathcal{G}^B} q_g - d_B = -f \tag{2.18c}$$

$$(\mu_g) \quad q_g \leq Q_g^{max} \qquad \forall g \in \mathcal{G} \tag{2.18d}$$

$$(\nu_i) \quad d_i \leq D_i \qquad \forall i \in \mathcal{N} \tag{2.18e}$$

$$(\overline{\kappa}) \quad f \leq \overline{f} \tag{2.18f}$$

$$(\underline{\kappa}) \quad f \geq -\overline{f} \tag{2.18g}$$

The optimality conditions of the problem are:

$$0 \leq q_g \quad \perp \quad MC_g - \pi_{i(g)} + \mu_g \geq 0 \qquad \forall g \in \mathcal{G} \tag{2.19a}$$

16 Let us notice that variants of this linear model exist (e.g. adding bounds on the angle differences). See Taylor (2015) for an extensive discussion of the topic.

17 More generally, this holds true for radial networks.

$$0 \leq d_i \quad \perp \quad \pi_i - VOLL + \nu_i \geq 0 \qquad \forall i \in \mathcal{N} \qquad (2.19b)$$

$$0 \leq Q_g^{max} - q_g \quad \perp \quad \mu_g \geq 0 \qquad \forall g \in \mathcal{G} \qquad (2.19c)$$

$$0 \leq D_i - d_i \quad \perp \quad \nu_i \geq 0 \qquad \forall i \in \mathcal{N} \qquad (2.19d)$$

$$0 = \sum_{g \in \mathcal{G}_i} q_g - d_A \pm f, \quad \pi_i \text{ free} \qquad \forall i \in \mathcal{N} \qquad (2.19e)$$

$$f \text{ free }, \quad \pi_B - \pi_A - \overline{\kappa} + \underline{\kappa} = 0 \qquad (2.19f)$$

$$0 \leq \overline{f} - f \quad \perp \quad \overline{\kappa} \geq 0 \qquad (2.19g)$$

$$0 \leq \overline{f} + f \quad \perp \quad \underline{\kappa} \geq 0 \qquad (2.19h)$$

Let us emphasize three main conclusions from equations (2.19). Firstly, the reader may check that these conditions are equivalent to the optimality conditions of consumers, suppliers, system operator and the market clearing constraints. Therefore, the fundamental efficiency property of a decentralized market discussed in the previous section still holds. The main novelty with respect to the previous section is the presence of a new agent, the system operator. In the model, this agent is assumed to behave as price-taker and to maximize his profit[18], the so-called *congestion revenue*, which comes from the the transmission capacity that is sold on the market. Mathematically, this is $\max\{f(\pi_B - \pi_A) | f \in \mathcal{F}\}$, with $\mathcal{F} = \{f \in \mathbb{R} | -\overline{f} \leq f \leq \overline{f}\}$ in the case of model (2.18). Secondly, marginal pricing still prevails in the market: at each node, the marginal supplier sets the price. Thirdly and most importantly, the optimality conditions also inform us on the relationship between locational prices. In case the line is not congested ($-\overline{f} < f < \overline{f}$), the prices in A and B are the same ($\pi_A = \pi_B$): as the grid resource is not scarce, the price associated to its usage is zero. In case the line is congested, say from A to B ($f = \overline{f}$), then the prices between A and B differ such that: $\pi_B = \pi_A + \overline{\kappa} \geq \pi_A$. Thus, the energy flows from less expensive to more expensive locations.

This fundamental behaviour is relevant for the discussion of chapter 3. Indeed, we shall see that under non-convexities in the production processes, although the network model remains linear, some of the pricing approaches do not necessarily preserve this sound relationship between locational prices: there may be—under *some* pricing rules—a price difference between two nodes in a radial network, although there is no congestion on the line connecting the two nodes[19].

18 The system operator being a monopoly seller of transmission capacity, regulation is needed to impede him from withholding capacity from the market.

19 Let us notice that, in general, this may also happen in a meshed network, even without non-convexities. But this does not invalidate the point we make in this paragraph; simply, the "sound relationship between locational prices" takes a slightly more sophisticated form in

2.6.3 *Competitive equilibrium with reserve*

In our production model, the two main commodities sold by a producer are energy and reserve. Since selling a MW of reserve prevents the producer from selling it as energy, the two commodities are tightly linked by arbitrage. The cost incurred by a supplier for providing an extra MWh of energy is determined by its marginal cost. The cost of providing reserve, on the other hand, is not linked to a direct operating cost, at least in our model. Instead, the cost of reserve is an *opportunity cost* of not producing energy. A unit that is out-of-the-money in the energy market would be ready to reserve capacity for 0€/MWh while a unit that is in-the-money, earning a profit of 10€/MWh in the energy market, would reserve its capacity for 10€/MWh.

Compared to model (2.9), the two modifications are the additional market-clearing constraint for reserve (2.20a) and the amendment of constraint (2.9c) to (2.20b) to include the provision of reserve:

$$(\pi_t^R) \quad \sum_{g \in \mathcal{G}} r_{g,t} \geq R_t \qquad \forall t \in \mathcal{T} \tag{2.20a}$$

$$(\mu_{g,t}) \quad q_{g,t} + r_{g,t} \leq Q_g^{max} \qquad \forall g \in \mathcal{G}, t \in \mathcal{T} \tag{2.20b}$$

The optimality conditions of the augmented problem are conditions (2.10) with the modification of (2.10d) to (2.21a) and the addition of (2.21b)–(2.21c). We assume here an inelastic demand of reserve R_t set by the system operator.

$$0 \leq Q_g^{max} - q_{g,t} - r_{g,t} \quad \perp \quad \mu_{g,t} \geq 0 \qquad \forall g \in \mathcal{G}, t \in \mathcal{T} \tag{2.21a}$$

$$0 \leq r_{g,t} \quad \perp \quad \mu_{g,t} - \pi_t^R \geq 0 \qquad \forall g \in \mathcal{G}, t \in \mathcal{T} \tag{2.21b}$$

$$0 \leq \sum_{g \in \mathcal{G}} r_{g,t} - R_t \quad \perp \quad \pi_t^R \geq 0 \qquad \forall t \in \mathcal{T} \tag{2.21c}$$

The reader may check that the decentralized interpretation of this augmented model still holds, as for model (2.9). The main conclusion regards the linkage between energy and reserve prices. Consider the marginal unit, that provides the system with both energy and reserve at time t ($q_{g,t}, r_{g,t} > 0$). The optimality conditions 2.10a, 2.21a and 2.21b imply that $\pi_t^R = \pi_t^E - MC_g$. This is exactly the arbitrage condition described above between selling energy or reserve.

This short analysis of reserve is introduced here for comprehensiveness and to demonstrate how it can be included in the model for future work.

a meshed grid (Papavasiliou, 2024; Taylor, 2015), and, in presence of non-convexities, *some* pricing rules may break this relationship.

Nevertheless, let us emphasize that we will not consider reserve in the remainder of this thesis. This is a limit of our analysis, although we do not expect the introduction of reserve in the model to fundamentally change the main conclusions and the results of the next chapters.

2.6.4 *Long-term competitive equilibrium*

LONG-TERM MODEL. As chapter 5 of this thesis studies the long-term effect of indivisibilities in investment decisions, it is worth outlining here the analysis of a *long-term* competitive equilibrium. Compared to model (2.9), which is a short-term model that assumes that suppliers operate with a given investment, a long-term analysis considers all the production factors as variable. Concretely, the capacity of the suppliers Q_g^{max} in model (2.9) is replaced by decision variables x_g that stand for the investment made in technology g, which comes at a cost IC_g. The model becomes:

$$\max_{q,x,d \geq 0} \sum_{t \in \mathcal{T}} \Delta T_t VOLL d_t - \sum_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}} \Delta T_t MC_g q_{g,t} - \sum_{g \in \mathcal{G}} IC_g x_g \qquad (2.22a)$$

$$(\Delta T_t \pi_t) \quad d_t \leq \sum_{g \in \mathcal{G}} q_{g,t} \quad \forall t \in \mathcal{T} \qquad (2.22b)$$

$$(\Delta T_t \mu_{g,t}) \quad q_{g,t} \leq x_g \quad \forall g \in \mathcal{G},\ t \in \mathcal{T} \qquad (2.22c)$$

$$(\Delta T_t \nu_t) \quad d_t \leq D_t \quad \forall t \in \mathcal{T} \qquad (2.22d)$$

While the set $\mathcal{T}$ in model (2.9) stands for a short-term set of periods (e.g. the set of hours of the next day), in model (2.22) $\mathcal{T}$ should be viewed as, for instance, the periods of an entire year. ΔT_t stands for the duration of period t. The optimality conditions of problem (2.22) are:

$$0 \leq q_{g,t} \quad \perp \quad MC_g - \pi_t + \mu_{g,t} \geq 0 \qquad \forall g \in \mathcal{G},\ t \in \mathcal{T} \qquad (2.23a)$$

$$0 \leq x_g \quad \perp \quad IC_g - \sum_{t \in \mathcal{T}} \Delta T_t \mu_{g,t} \geq 0 \qquad \forall g \in \mathcal{G} \qquad (2.23b)$$

$$0 \leq d_t \quad \perp \quad -VOLL + \pi_t + \nu_t \geq 0 \qquad \forall t \in \mathcal{T} \qquad (2.23c)$$

$$0 \leq x_g - q_{g,t} \quad \perp \quad \mu_{g,t} \geq 0 \qquad \forall g \in \mathcal{G},\ t \in \mathcal{T} \qquad (2.23d)$$

$$0 \leq D_t - d_t \quad \perp \quad \nu_t \geq 0 \qquad \forall t \in \mathcal{T} \qquad (2.23e)$$

$$0 \leq \sum_{g \in \mathcal{G}} q_{g,t} - d_t \quad \perp \quad \pi_t \geq 0 \qquad \forall t \in \mathcal{T} \qquad (2.23f)$$

These equations, which formalize the "peak-load pricing" analysis pioneered by Boiteux (1960), convey three important facts. Firstly, as previously, model (2.22) admits a decentralized interpretation, so that marginal pricing provides market agents with the right incentives to invest in the surplus-maximizing generation mix. Secondly, if a technology is used ($x_g > 0$), then

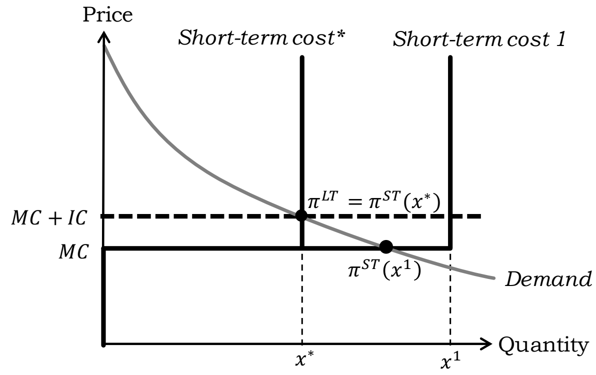


FIGURE 2.3: Analysis of an investment problem with one technology and a uniform demand along the year.

the infra-marginal rents ($\sum_{t \in \mathcal{T}} \Delta T_t \mu_{g,t}$) earned from the short-term market prices π_t by each technology exactly cover the investment cost IC_g. This means that long-term profits are zero. Thirdly, as highlighted by Boiteux, in order for the peaking units (the technology g with the highest MC_g) to recover their fixed costs, there should be at least some hours during which the demand sets the price higher than marginal cost ($d_t \leq D_t$ and $\nu_t \geq 0$ such that $\pi_t = VOLL - \nu_t > MC_{peak}$).

GRAPHICAL ILLUSTRATION. For the sake of illustration, let us consider the case of a single technology and a demand that is uniform along the year. Figure 2.3 illustrates this case. In the short-term, as in model (2.9), the investment x is fixed and the investment cost IC is sunk. The market clearing price reflects the *short-term marginal cost* of production. In case the production in the short-term is such that $q < x$, as for the (sub-optimum) investment x_1 in Figure 2.3, then the market clearing price is $\pi^{ST} = MC$. Under this configuration (over-investment), the supplier does not recover its investment cost. Therefore, he would reduce his capacity x until the investment is recovered. In case the production of the supplier is saturated ($q = x$), the price is set by the demand such that $\pi^{ST} = MC + \mu$ for a certain μ. Which value of μ would the investment x face in equilibrium? The answer provided by model (2.22) is $\mu = IC$ such that $\pi^{LT} = MC + IC$. At the optimum investment x^* on Figure 2.3, the market-clearing price equals both the *short-term* and *long-term marginal cost*: $\pi^{ST} = \pi^{LT} = MC + IC$.

RELATIONSHIP WITH NETWORK EXPANSION. Chapter 5 discusses the problem of optimal investment *in power generation assets* (in engineering term the "generation capacity expansion problem"). But it does not include the second main problem of investment encountered in the electricity market, namely investment in power lines (or the "transmission planning problem"): the problem faced by the system operator who invests in new line capacity. In other words, our analysis will focus on the generation investment problem—and in particular the effect of lumpy investments—assuming a fixed network topology. Although we do not further treat the subject in the thesis, let us conclude this section by highlighting the linkage between the two problems of *transmission capacity investment* and *generation capacity investment*.

From an intuitive point of view, the value of an investment in a line between two nodes is essentially the value of replacing a costly generation source in one of the nodes by another cheaper one in the other node. Let us consider again the two-node model (2.18) of section 2.6.2. Let us neglect the multi-stage nature of the problem and assume convex investment decisions in the line capacity simultaneously with convex investment in generation assets. Compared to model (2.18), the novelty is the introduction of investment variables for the suppliers (x_g in place of Q_g^{max}) and for the network (y in place of $\overline{f}$). The investment in the line comes with a cost $LC \cdot y$ that appears in the objective function. The investment in generation asset g also appears in the objective function, as in model (2.22), with cost $x_g IC_g$. Then, omitting consumer equations and market-clearing constraints, the optimality conditions are:

$$0 \leq q_g \;\perp\; MC_g - \pi_{i(g)} + \mu_g \geq 0 \qquad \forall g \in \mathcal{G} \tag{2.24a}$$

$$0 \leq x_g - q_g \;\perp\; \mu_g \geq 0 \qquad \forall g \in \mathcal{G} \tag{2.24b}$$

$$0 \leq x_g \;\perp\; IC_g - \mu_g \geq 0 \qquad \forall g \in \mathcal{G} \tag{2.24c}$$

$$0 \leq y \;\perp\; LC - \overline{\kappa} - \underline{\kappa} \geq 0 \tag{2.24d}$$

$$0 \leq y - f \;\perp\; \overline{\kappa} \geq 0 \tag{2.24e}$$

$$0 \leq y + f \;\perp\; \underline{\kappa} \geq 0 \tag{2.24f}$$

$$\pi_A - \pi_B + \overline{\kappa} - \underline{\kappa} = 0 \tag{2.24g}$$

Let us assume that the generation assets are cheaper in node A such that, at the optimum, energy flows from A to B ($y = f > 0$ and $\underline{\kappa} = 0$). Then $LC = \pi^B - \pi^A$, with $\pi^A = MC_{g_A} + IC_{g_A}$ (for a given marginal unit $g_A \in \mathcal{G}_A$) and $\pi^B = MC_{g_B} + IC_{g_B}$ (for a given marginal unit $g_B \in \mathcal{G}_B$). This says that, at the optimum, the extra cost of the investment in the line capacity equals the value gained by supplanting production in B by production in A. Suppose that this is not the case and that, instead $LC < \pi^B - \pi^A$. Then, investing in one more MW of line capacity would cost LC while it would save $\pi^B - \pi^A$ to the consumers, who can buy one more MW from the cheaper generation in A. Since $LC < \pi^B - \pi^A$, this would result in a net increase of total surplus, meaning that the solution would not be optimal.

2.7 MARKET ORGANIZATION AND ARCHITECTURE

So far, the analysis has abstracted from the effective organization of the marketplace. An exception has been the European market model of production (2.7), which examined a concrete model used by the European power exchanges in the actual day-ahead energy auction. Although the theme of this thesis, the price formation in a non-convex market, addresses a very theoretical and fundamental issue of how to determine a price in a non-convex market, it also has a practical dimension, as pointed out in the Introduction. Since electricity markets are typically organized as closed-gate auctions involving non-convex bids, the rules that should be adopted by the auctioneer to clear the market and to compute prices are not merely a theoretical issue but a very practical one. This becomes a *market design* issue, or a *market architecture* issue, in which one should define what are the "rules of the game" (Stoft, 2002).

According to Coase, "markets are institutions that exist to facilitate exchange" (Coase, 1988, p. 7). Acknowledging their *institutional* nature means stressing the importance of defining rules and regulations that permit these "market institutions" to work properly.

> "All exchanges regulate in great detail the activities of those who trade in these markets (the times at which transactions can be made, what can be traded, the responsibilities of the parties, the terms of settlement, etc.), and they all provide machinery for the settlement of disputes and impose sanctions against those who infringe the rules of the exchange. It is not without significance that these exchanges, often used by economists as examples of a perfect market and perfect competition, are markets in which transactions are highly regulated (and this quite apart from any government regulation that there may be). It suggests, I think correctly, that for anything approaching perfect competition to exist, an intricate system of rules and regulations would normally be needed." (Coase, 1988, p. 9)

This section outlines the "intricate system of rules and regulations" that have been adopted in the electricity sector in the pursuit of "approaching perfect competition", the ideal described in section 2.6. Our discussion is structured around three main characteristics, that could appear as peculiarities of the electricity sector: (i) the highly *centralized* way in which electricity markets are organized, (ii) the *multi-settlement* nature of the market and (iii) the *multi-product* nature of the market.

2.7.1 *Centralized market*

From the discussion of section 2.3, one may conclude that a minimum level of centralization is needed in power markets. Even decentralized bilateral trades of energy would ultimately need a central authority for the auctioning of transmission capacity, to control the provision of ancillary services, or to take care of measurements. Nonetheless, although it is generally acknowledged that some centralization is needed, the questions of how much the market should be centralized as well as the scope of the system operator have been subject to debates.

MARKET TYPES. There are multiple ways to arrange electricity market trades. From the least to the most centralized, these include (Stoft, 2002): (i) *bilateral* market trade, only involving two parties or (ii) *mediated* market trade, which involves a third party arranging the trade. This can be a dealer (a third party holds the commodity before reselling it and takes the spread) or exchanges and pools, typically organised as auctions. In the taxonomy of Stoft, pools are distinct from exchanges essentially by the presence of side-payments and complex bidding structure. More centralization implies a less flexible market (or more standardized products) but faster—or less costly—transactions and a higher level of coordination. The bilateral trade minimizes the role of the system operator, while the pool maximizes it. From the perspective of bidding complexity, if the model behind the the pool is a unit commitment model, then it minimizes the bidding complexity for the traders since they merely have to submit their cost structure. The bilateral trade greatly increases the complexity since the traders would have to trade with different locations and acquire explicit transmission rights separately. The exchange may be viewed as a middle ground.

The discussion about the degree of centralization that should be adopted in electricity markets has received various namings in the literature. The distinction has sometimes being phrased as between "bilateral trade", "exchange" and "pool" (Stoft, 2002), or between "integrated' *vs* "unbundled system" (Wilson, 2002) or between "integrated market" and "exchange-based market" (Cramton, 2017). I would rather emphasize multiple choices that have to be made in the market, which are to some extend independent, and characterize the degree of centralization of the market:

- *Centralized vs decentralized market for energy*: most US day-ahead markets today, as well as the EU day-ahead market, are centralized.

California, in the early 2000s, used to rely on a decentralized market for energy.[20]

- *Integration of energy, transmission and reserve*: to what extent does the market integrate these three commodities and optimize their allocation jointly.
  - Nodal *vs* zonal: the US markets, organized by the different US ISOs, are nodal while the European market is zonal. The fact that EU is zonal means that a lot of transmission is not allocated by market mechanisms but by out-of-market—administrative—procedures.
  - Co-optimization of reserve and energy: some US markets perform co-optimization, the European market does not (although it cooptimizes to a certain extent energy and transmission, which invites the natural question of why one leaves reserve out of the picture).
- *Convexity of the market model*: non-convex markets permit a refined scheduling model that accounts for inter-temporal constraints of operations, cost structure, and so on. That is, it enables the market to provide a higher degree of *coordination* between suppliers—coordination that the main attributes of electricity, discussed in section 2.1, are calling for. But it also creates new challenges, especially price formation issues that are the subject matter of this thesis.
  - Convex *vs* non-convex market: many US markets rely on a unit commitment model which is non-convex. Although the European day-ahead market does not rely on unit commitment, it is a non-convex market. New-Zealand is an example of a convex day-ahead auction for power (Bergheimer et al., 2023).
  - Inclusion of side-payments: the US markets pay side payments (the definition of side payments may in itself vary: some ISOs pay “make-whole payments” while ISO-NE pays the so-called “lost opportunity costs”, cf. chapter 3). The European market, although it is non-convex, does not pay any side payments.

The chosen level of centralization may of course depend on the time frame of the market. It is admitted that the real-time market for power

20 California then experienced one of the biggest failure in electricity markets history, which arguably throw some doubts on the effectiveness of such a decentralized market design. As Hogan puts it: “California built its market design on a flawed premise that the inescapable reality of coordination requirements could be ignored or minimized in an effort to honor a boundless faith in the ability of markets to solve all problems.” (Hogan, 2002)

should be organised in a highly centralized fashion, since a high level of coordination is required and has to be achieved in a few seconds. On the other hand, a long-term forward market allows more room for the possibility of more decentralized market types. The organization of the day-ahead market has been more controversial. Stoft (2002) tends to argue for a centralized nodal market on the ground that it does render the computations of the system operator harder while enhancing the optimization of power flows. But he rather argues for simpler convex auction which avoid the dilemma of side-payments. Cramton (2017) revisits this debate, and rather argues for a more integrated model which is centralized, nodal and relies on unit commitment in day-ahead and stochastic economic dispatch in the real-time market. His arguments rely on the experience over the past decades, the improvements in software and optimization algorithms and, most importantly, the fact that centralized nodal and non-convex auctions can ultimately do a better job in accounting for both power unit constraints as well as transmission constraints, resulting in a more efficient usage of resources and sound price formation that reflects more accurately the scarcity of the system.

AUCTION MODEL. Centralized electricity markets are typically organised as closed-gate (one-sided or two-sided) sealed-bid uniform price auctions. However, they differ greatly in their actual functioning rules. These include:

- *The bidding rules*: the format of the bids accepted by the auction. These may be "unit" types of bids as in the unit commitment model (2.5) in which bidders submit plant technical characteristics, or European "portfolio" types of bids as in model (2.7).
- *The bid acceptance rules*: the rules determining which bids are cleared. This can be a surplus maximization model, or more complex rules as in the European day-ahead market.
- *The price formation rules*, which are of particular importance for the subject matter of this thesis. These rules vary largely in the presence of non-convexities in the electricity auctions. These will be discussed in detail in chapter 3.
- *The settlement rules*, mainly, whether the uniform price is complemented by discriminatory side-payments or not, and if so, how these payments are allocated (i.e. both payments *to* and payments *from*).

Let us notice that it is mainly the "bid acceptance rules" that play a key role in determining the total surplus of the market. The price formation rules affect the *split* of this surplus between agents. Even in case the price formation rules vary in the amount of side-payments that have to be paid, to the extent that these side-payments are financed through the surplus of other agents in the market, this again does not affect the total surplus but merely its distribution (though it may affect incentives for bidding and thus actual allocations).

PAY-AS-BID *VS* UNIFORM PRICE AUCTION. An important notion for this thesis is that of a *uniform* price auction, in which the cleared bids receive the same price. An alternative is *pay-as-bid* auctions, in which the price received by each cleared bid is its own bid price. Of course, a pay-as-bid auction would render the issue of non-convexities less problematic, since the auctioneer does not have to find a uniform price that is applied to all transactions, but instead compensate each bid at the bid price. Nonetheless, a pay-as-bid auction creates other inconveniences, which justify why most day-ahead electricity auctions have adopted uniform pricing approaches[21].

In a pay-as-bid setting, the consumer bill is expected to be reduced *if all agents bid truthfully*. This was the naïve expectation advocated for pay-as-bid in the California market (Kahn et al., 2001): the pay-as-bid scheme would mitigate the frequent price spikes. Nonetheless, in the pay-as-bid settings, producers have incentives to deviate from marginal cost bidding: the approach is not *incentive compatible*. The outcome may be equivalent to uniform-price auctions under certain idealized assumptions (cf. revenue equivalence theorem). However, in practice the design can result in inefficiencies, due to the fact that bidders will tend not to bid truthfully: bidders are incentivised to guess the uniform price outcome and forecasting errors may lead to welfare-enhancing bids being rejected. Furthermore, bidders would incur additional costs coming from their new pressing need to forecast the market price. This may in turn alter competition: as profits become related to the ability of a firm to forecast, small competitors can be disadvantaged compared to bigger players. Ultimately, if pay-as-bid were successful in holding the price below the competitive level, this would distort long-term investment incentives, as the infra-marginal rents earned by the suppliers in the market are needed to recover investment costs (cf. section 2.6).

2.7.2 *Multi-settlement market*

SPOT AND FORWARD MARKETS. Electricity is finally exchanged in real time between producers and consumers through the electrical grid. How-

21 For a discussion on pay-as-bid, see Stoft (2002), and also Kahn et al. (2001).

ever, in order for this to happen, there is a sequence of markets that take place from far ahead in the past until the real-time market. More specifically, two cornerstones of an effective electricity market are the *day-ahead* and the *real-time* markets. There also exist longer-term forward markets.

The real-time market is the spot—*physical*—market: a physical commodity is sold "on the spot" and the trades correspond to true power exchanges. All other markets, including the day-ahead market, are forward—*financial*—markets: there is no delivery of a physical commodity but only trades of financial contracts, which are derivatives of the real-time price of electricity. A central notion to keep in mind when apprehending this sequence of markets is the direction in which the sequence should be read (Hogan, 2022). Against the "engineering logic" which tends to move *forward*, starting from a rough planning in weeks or days in advance towards the refined scheduling in real-time operations, the "market logic" moves *backward*: it starts from the real-time expected conditions and price, and the incentives these imply for the the market participants, and it back-propagates to forward markets. Therefore, as argued by Hogan, a sound market design should start fixing the real-time spot price of power and work backward to the forward markets, and not the other way around.

The separation of physical and financial markets is arguably a feature of a successful electricity market.[22] Acknowledging this distinction also explains why several day-ahead markets, such as ERCOT, include financial *virtual bidding* to foster price discovery—i.e. the convergence of real-time and day-ahead prices (Hogan, 2016).

TWO-SETTLEMENT PRINCIPLE. This sequence of markets implies a sequence of settlements. The *two-settlement* principle according to which the day-ahead market and the real-time market are interconnected may be described as follows. The day-ahead market trades *forward contracts*, in which the seller is paid the day-ahead market for the quantity sold and commits to deliver energy or to buy back his position at the real-time market price when realized. Thus, if a seller sells q_{DA} at the DA price π_{DA}, and then delivers q_{RT} in real time, he is paid: $q_{DA}\pi_{DA} + (q_{RT} - q_{DA})\pi_{RT}$. This two-settlement system preserves the incentives of the market participants in the real-time market. The above seller bids in the real-time market as if nothing has been sold in the day-ahead market. Indeed, his profit maximization objective in real-time is: $q_{DA}(\pi_{DA} - \pi_{RT}) + q_{RT}\pi_{RT} - C(q_{RT})$. Assuming the seller is a price-taker, the first term is a constant *sunk cost*, which implies that the profit-maximizing solution in real time is independent of the

22 As Hogan (2016) phrases it: "An important feature of successful electricity market design is the necessity to separate the financial role of contracts used to allocate risk and the physical operation of the system."

day-ahead transactions. As far as the settlement is concerned, from the above formula, the electricity delivered in real time can be split between the share already sold under a forward market contract, paid at the forward price, and the share sold in real time at the real-time price, which is the so-called *imbalance* with respect to the position in the forward market.

THE DAY-AHEAD MARKET "The day-ahead market is the forward market with the greatest physical implications" (Stoft, 2002, p. 243). Indeed, although it is a financial market, it plays a crucial role in deciding the actual operational schedule, and providing the unit commitment "service" to the market players. This explains why the day-ahead market often includes a complex bidding structure, in order to account for the many technical constraints of the scheduling of power plants. This market will be central for this thesis, as it turns out to be the power market *par excellence* that includes non-convexities.

As mentioned earlier in section 2.7.1, the organization of the day-ahead market has been particularly controversial, such that, in practice, there exists a myriad of rules and specificities. A bilateral trade for power combined with a centralized market for transmission rights would be a possible option for the day-ahead market. This would minimize the role of the system operator. However, reaching an *efficient* equilibrium may be difficult because of the decentralized nature of the approach. In particular, the separation of the transmission market and the energy market makes the bidding task more complex for the traders. This may harm both the efficiency of the allocation, as well as the reliability of the system. In practice, day-ahead markets have adopted a more centralized organization. To leverage the taxonomy introduced in section 2.7.1, the day-ahead markets we will study in this thesis include non-convexities and coordinate the auctioning of energy and transmission capacity. The US day-ahead market held by PJM, ERCOT or NYISO relies on unit commitment models, similar—but more sophisticated—than model (2.5). Instead, the European day-ahead market relies on model (2.7).

This thesis starts from the presence of non-convexities in various day-ahead electricity auctions and studies the problem of determining the price in these circumstances. However, it is fair to mention here two arguments against the mere idea of including non-convexities in the auction in the first place. (i) First, while a unit commitment model theoretically minimizes all the costs with a great level of technical detail, it assumes that agents will reveal their costs and parameters truthfully. As Wilson (2002) puts it: "absent regulatory enforcement, cost minimization is a fiction without

	2016	2017	2018	2019	2020	2021	2022	2023	2016–2023
DA av. price	36.6	44.6	55.3	39.3	31.9	104.1	244.5	97.3	81.7
DA std	23.5	21.6	23.5	18.0	16.5	79.4	134.7	45.9	89.7
RT av. price	34.9	42.2	53.4	39.1	33.8	100.3	233.6	96.7	79.2
RT std	46.9	56.3	66.4	51.0	54.2	129.8	222.4	143.2	129.6
Abs. difference	−1.7	−2.4	−1.9	−0.2	1.9	−3.8	−10.9	−0.6	−2.4
Rel. difference	−5%	−6%	−4%	−1%	6%	−4%	−5%	−1%	−3%

TABLE 2.3: Price convergence between day-ahead and real-time markets in Belgium. All figures are in €/MWh. [Data source: ENTSO-E Transparency Platform (DA price) & Elia website (RT price, i.e. 15-min imbalance price)]

stronger incentives to ensure that bids reflect actual costs."[23] (ii) Second, the day-ahead prices are forwards of the expected real-time prices. Table 2.3 illustrates the convergence of day-ahead and real-time prices in Belgium over the last years[24]. The day-ahead market is a forward market, but it is special since the price is not the result of an *arbitrage* by the agents but of a *computation* made by the auctioneer. Yet, since the day-ahead market has fixed rules and variable inputs, it may be that the inputs are actually determined by the market participants as expressing an arbitrage with the expected real-time price. If this were the case, the complex bidding rules and price formation rules would not increase coordination but simply render the arbitrage more complicated (Stoft, 2002).

THE REAL-TIME MARKET The real-time market is operated in a highly centralized fashion, essentially because time is an issue: the system operator needs to balance the system continuously, to ensure security of transmission and system reliability, and this could not be achieved by a decentralized market. The real-time market is often even more than a "centralized market" that relies on the price signal to coordinate resources: it typically involves *direct* interventions and activations of resources by the system operator, as "quicker coordination than the market provides" is needed (Wilson, 2002). This explains why this market is often termed a "balancing market", at least in Europe. In Belgium, these so-called "balancing activations" are not based on a centralized market clearing tool or auction, as in the day-ahead market, but are performed by the system operator, and follows some

23 Let us notice that US ISOs typically implement some market power mitigation measures in the auction, such as caps on the bid's price.

24 As a point of comparison, the difference between the average real-time and day-ahead prices in PJM was -0.06 and 0.06$/MWh in 2017 and 2018 respectively (Hogan, 2021).

approximative merit-order rule, but without the non-convexities and subtle price formation rules that exist in the day-ahead market. This is however evolving towards a more market logic with the creation of pan-European balancing platforms.

2.7.3 *Multi-product market with multi-part bids*

As discussed in section 2.3, the delivery of electrical energy in real time to consumers requires the trading of multiple associated commodities. Therefore, an electricity market is often a multi-product market: not only does it trade energy on multiple periods and locations at the same time, but it also includes the auctioning of transmission rights, and sometimes reserve. As discussed in section 2.7.2, the focus of this thesis will be the day-ahead market, which typically includes at least energy for the 24 hours of the next day (in 1-hour or 15-minute resolution) as well as transmission rights. The price formation rules are therefore a multi-dimensional problem, involving the computation of tens, hundreds or even thousands of prices in a single market session (as an example, the biggest market session that is solved in chapter 4 involves the computation of more than 5500 prices). The prices of these various products are therefore tightly linked within one market. These linkages have been highlighted in section 2.6. The objective of this section is to comment on their *explicit* or *implicit* existence in the actual organisation of the market. In particular:

- *Temporal linkage*. The electricity auctions we will be studying, such as the day-ahead market, are multi-period auctions. Therefore, there is an *explicit* linkage of the prices of energy between the periods which results from inter-temporal constraints in the bids submitted to the market, as well as, occasionally, operational constraints of the network (such as ramp constraints on the lines, which are implemented in the European DA market (NEMO Committee, 2020b)). A temporal linkage also exists between different auctions that trade products for the same delivery period, although it remains *implicit*: a forward market price, such as the day-ahead market price, for tomorrow at noon would approximate the spot price of tomorrow at noon (cf. Table 2.3, although the real-time prices are typically much more volatile).

- *Spatial linkage*. The electricity auctions analysed in this thesis also include explicit spatial linkages between locations. This results from the simultaneous auctioning of energy and transmission capacity. However, the granularity of this spatial linkage varies. For instance, the price in France and the price in Germany are explicitly linked in the day-ahead EU market. Some other spatial linkages (for example between north and south Germany), *although they exist*, are ignored by the market. Since the EU market does not provide locational nodal price signals down to the level of a substation but merely a regional zonal price signal at the level of a country, some of the existing spatial linkages are ignored. Ignoring these linkages within the market creates, on the long-term, poor incentives for investing in intra-zonal lines, distortion of generation investments and, in the short-term, the activation of bids that violate the grid constraints. This in turn necessitate out-of-market intervention, such as redispatch[25], which in turns creates gaming opportunities that are foreseen in theory and observed in practice—that continues to plague Europe, and that contributed to the collapse of the California market in 2001.

- *Reserve–Energy linkage*. This linkage is *explicit* in some US auctions that co-optimize energy and reserve. In Europe, this linkage remains *implicit*, as the reserve and energy markets are separated. In this thesis, we do not model this linkage.

25 Multiple European initiatives try to solve this problem indirectly. Local energy markets that include intra-zonal grid constraints can be viewed as a way to overcome the challenges that a zonal market artificially creates. See for instance the discussion in Mezghani et al. (2023).

Part I

Non-Convexities in the *Short-Term*

3

ON SOME ADVANTAGES OF CONVEX HULL PRICING

ABSTRACT. *Since the liberalization of the power sector and the creation of wholesale electricity markets, the question of how to price the non-convexities that are present in the market has attracted the interest of both academics and practitioners. Over the years, US markets have studied and adopted different and evolving pricing rules. Since the "Trilateral Market Coupling" (2006), the European day-ahead market has opted for a notably different pricing rule. Recently, EU stakeholders have undertaken research to reform it, and have indicated an interest for some approaches that are discussed in the other side of the Atlantic. This chapter aims at contributing to the debate. We analyse six different pricing methods. We establish several mathematical properties for enabling their accurate comparison. Our findings are illustrated on stylized examples and numerical simulations that are performed on realistic datasets. Both theoretical and numerical evidences that are gathered in this chapter point towards the advantages of convex hull pricing*[*].

KEYWORDS. Convex hull pricing · Non-uniform pricing · Non-convexities · European electricity market

JEL CLASSIFICATION. C61 · D41 · D44 · D47 · Q41

* The chapter reproduces, with minor changes, the content of Stevens et al. (2024a).

3.1 INTRODUCTION

POWER auctions are notably characterized by the presence of non-convexities. In the US, these non-convexities emerge from the so-called unit commitment model, which has been run in control rooms since before the liberalization of the power sector took place[1]. Although some economists have argued for simpler—convex—market models (cf. the arguments covered by Stoft (2002), outlined in section 2.7.2), unit commitment has prevailed in many US auctions. In Europe, despite the fact that the market model is different, it also includes non-convex bids, the so-called "block orders" being the simplest example.[1] Although the European market does not rely on *physical* unit commitment models, the non-convex orders also aim—indirectly at least—at providing the suppliers with the flexibility of representing the complex constraints of power generation into the auction. Non-convex multi-parts bids are a bet that the efficiency gained by a refined scheduling model (and the improved coordination[2] between suppliers that this enables), are higher than the inefficiencies resulting from the increase in complexity. In particular, the main drawback of non-convexities is that they impede the existence of a competitive equilibrium. The "classical" marginal prices fail to support the efficient allocation of goods. The absence of equilibrium prices has resulted in various and evolving pricing practises among the US and EU markets.

The liberalization of the power sector in the US started in the 90s, encouraged by the government through the Energy Policy Act of 1992. The creation of the Independent System Operators (ISOs), that have assumed the role of operating the market, followed in the late 90s and early 2000s. Locational *marginal pricing* (LMP) has traditionally been adopted by many ISOs to clear the market, cf. Stoft (2002) and the historical account provided by EPRI (2019). Experience revealed several drawbacks of marginal pricing, especially the fact that short-term fixed costs are not reflected in the price signal which therefore does not provide adequate incentives to market participants. The inadequacy of marginal pricing has stimulated research about the right way to price non-convex power auctions. Convex hull pricing (CHP) (Hogan and Ring, 2003) has emerged as a promising—although contested (Schiro et al., 2015)—way to price energy in the presence of non-convex bids. Acknowledging these issues, several ISOs started moving away from marginal pricing. In 2014, the US Regulatory Commission launched a consultation about price formation in power auctions (FERC,

1 cf. section 2.5.
2 The need for coordination in power systems is justified by the main attributes of electricity, described in section 2.1.

2014). In 2015, MISO implemented "Extended LMP" (ELMP, an approximation of convex hull pricing) and a similar proposal followed by PJM in 2017 (PJM, 2017). Other ISOs have implemented various "fast-start pricing" approaches (EPRI, 2019), which are variants of ELMP. They typically share the property of including, to some extent, fixed costs in the price and resorting to some sort of linear relaxation of the problem for computing market clearing prices. One example is the "hybrid pricing" approach, or "Fixed Block Unit Pricing", implemented by NYISO (2016) since the early 2000s. That being said, up to recently, some ISOs such as CAISO or SPP still rely on marginal pricing (CAISO, 2020; EPRI, 2019).

The restructuring of the power sector in Europe went down a similar path, notwithstanding its peculiarities, cf. the historical account by Meeus (2020). Following the creation of the European Single Market in 1993, the First Energy Package initiated the liberalization of the power sector in 1996. The actual unbundling of competitive (supply and retail) and regulated (TSO and DSO) segments effectively took place between 2003 and 2009 (the Second and Third Energy Packages), along with the creation of national Regulatory Authorities. The implementation of power markets followed, with a different institutional arrangement than in the US: instead of the US ISOs (private, *non*-profit), the EU market is operated by the Nominated Electricity Market Operator (NEMO, private and *for*-profit). The first centralized—and non-convex—auction, coupling parts of central-western European countries, went live in 2006 (the so-called "Trilateral Market Coupling"). This auction has been progressively extended to more member states and in 2014 it became the Single Day-Ahead Coupling (SDAC) that still prevails today. SDAC currently couples 27 countries (62 bidding zones, 30 TSOs and 16 NEMOs) with an average daily traded volume of 4.62 TWh for a market surplus of 9.9B€ per session (NEMO Committee, 2023).

The pricing approach adopted early on by SDAC (NEMO Committee, 2020b), inherited from the design of the Trilateral Market Coupling (Belpex et al., 2006), significantly differs from those encountered across the US. A central difference in the design is the introduction of side-payments. Because an equilibrium does not exist with a uniform energy price, the US ISOs resort to discriminatory side-payments that complement the uniform price of energy. In contrast with this—so-called in EU parlance—"*non-uniform* pricing", the EU stakeholders have opted for a *uniform* pricing rule. This is anchored in the regulation: the Market Codes emphasize the importance for the payments to be non-discriminatory (CACM GL, Art. 38, 1.b, cf. Commission Regulation (EU) (2015)). According to Meeus (2020), this implies that the introduction of "non-uniform pricing" (i.e. the usage of side payments) would require to change the regulation. This has motivated market clearing rules that are notably different from those in the US. The general principle of the EU pricing approach can be described as follows. It is deemed unacceptable for a non-convex bid, such as block orders, to be cleared while it is out of the money (a so-called "paradoxically accepted block", or PAB). Since the market principles reject the usage of side payments, the market may not clear PABs. Thus, the auction first solves the dispatch problem by aiming at maximizing the welfare. Then, if no price can be found that respects the no-PAB requirement, some constraints are added to the dispatch problem which is solved again. This process repeats until the set of allocation and price satisfies all the requirements[3].

There are three main issues with this pricing approach (Van Vyve, 2011). Firstly, as opposed to US auctions that clear the welfare-maximizing allocation, the EU market clearing rules can result in rejecting welfare-enhancing bids in order to satisfy the no-PAB requirement. From an economic viewpoint, this welfare loss is critical since *efficiency* (maximization of the total surplus) is the main justification for the market to exist[4]. From a regulatory standpoint, the CACM GL market codes (Art. 38, 1.a, cf. Commission Regulation (EU) (2015)) specifically emphasise that the EU pricing algorithm should "aim at maximising economic surplus for single day-ahead coupling", which is, strictly speaking, currently not the case. Secondly, although the EU pricing rule ensures no PAB orders, the outcome is *not* a competitive equilibrium. There are market participants that are not cleared while they would be profitable: the so-called "paradoxically rejected blocks" (PRB). In 2022, there was an average volume of 12GWh of PRBs per bidding zone per day, which amounted to a total profit loss of 129 thousand euros per day (NEMO Committee, 2023). From a regulatory viewpoint, using the previously cited Art. 38 1.b of CACM GL, one could argue that the current pricing rule *already* entails discrimination of market players through the PRBs. Thirdly, the complexity of the clearing rules creates computational challenges. This is problematic, since the current algorithm is granted 17

3 To simplify the exposition, we only describe the PAB requirement. As a matter of fact, there are additional "primal-dual" constraints in the market rules, that an interested reader can find in NEMO Committee (2020b).
To provide the reader with more intuition, this pricing rule could be further illustrated with the stylized example of Figure 1.1, developed in the introduction of the thesis. The European pricing rule implies rejecting supply bid S_1, and, instead, clears bids S_2 and D_1. Assuming $Q(D_1) > Q(S_2)$, the price would be $P(D_1)$. As one may observe, the welfare loss implied by this rule is critical in this example.

4 Unfortunately, there is no public figure regarding the welfare loss in the European day-ahead auction, although it is a key indicator. ACER is the institution that defines the KPIs that are reported in the annual CACM reports. It would arguably make sense to include this additional KPI: the difference of welfare between the "root node" of the market clearing algorithm and the final solution.

minutes to compute the market clearing allocation and price for the entire European continent. This limit increased from 12 to 17 minutes between 2019 and 2022—and there are discussions to further extend it to 30 minutes or more (MCSC, 2023)—, reflecting the computational stress caused by this pricing requirement. The Market Codes also emphasize the importance of "scalability", cf. CACM GL, Art. 38, 1.e in Commission Regulation (EU) (2015).

For these reasons, SDAC is undertaking research to reform the current pricing rule (SDAC, 2023). Initial EU stakeholder discussions on "non-uniform prices" identified convex hull pricing as one possible option for the EU market (NEMO Committee, 2020a). More recent discussions have rather focused on marginal pricing (MCSC, 2022), although nothing is decided yet (SDAC, 2023). This chapter aims at contributing to these discussions relative to the reform of the European pricing rules, although our analysis also applies to US auctions. Our discussion focuses on possible alternatives to the current pricing rule, i.e. we discuss the advantages of these alternatives *between* them and not *over* the SDAC pricing rule. In particular, the contributions of this chapter are threefold.

Firstly, we perform a cross-comparison of four different pricing approaches. Several properties are formalized mathematically on the same model, in order to allow for a rigorous comparison of the alternative prices. This chapter focuses on the *short-term* properties of the prices. The *long-term* properties—the effect of pricing on investment incentives—have notably been studied in other recent works (Mays et al., 2021; Byers and Hug, 2023). Our endeavor aims at addressing the urge for a better understanding of various pricing candidates, as called upon by EPRI (2019). To some extent, we follow up on the pioneering works of Schiro et al. (2015) and Liberopoulos and Andrianesis (2016). While Schiro et al. (2015) focus solely on Convex Hull Pricing, we discuss the later in comparison with other approaches to better grasp their relative benefits and drawbacks. We also critically review some of the arguments provided by Schiro et al. (2015). While Liberopoulos and Andrianesis (2016) study some properties on a "two-suppliers" model, we rather analyse other properties on a general market model.

Secondly, the theoretical properties are supported by numerical simulations on realistic systems. This is a novelty compared to both Schiro et al. (2015) and Liberopoulos and Andrianesis (2016). In particular, studying convex hull pricing on realistic instances is an effort that has not been widely undertaken in the literature. Thanks to recent algorithmic progresses (Stevens and Papavasiliou, 2022; Andrianesis et al., 2021) we are able to compute *exact* CHP on realistic instances. This enables an accurate numerical comparison. More specifically, we illustrate and study the properties of the four pricing approaches on two different datasets: the "FERC dataset" (public data, but without a network) and the "CWE dataset" (non-public data, but including a network).

Finally, we particularly include the pricing method proposed by Madani and Papavasiliou (2022) referred to as "Minimal Make-Whole Payment" (MMWP) pricing in our comparison. This novel approach is representative of various recent proposals that have appeared in the literature, which have not been critically assessed so far. We notably implement three alternative versions of MMWP, and we discuss their relative advantages.

The material of the chapter is organized as follows. Sections 3.2 and 3.3 introduce the model, the main concepts and the four pricing schemes. Sections 3.4 to 3.8 then study their properties, and provide results from numerical simulations. To some extent, sections 3.5, 3.6 and 3.7-3.8 focus respectively on the comparison between CHP vs MMWP, CHP vs ELMP and CHP vs marginal pricing.

3.2 MARKET MODEL AND DISTANCE TO EQUILIBRIUM

Throughout this chapter, we consider the following auction model, which can accommodate the settings of both the EU day-ahead market[5] as well as most US auctions.

$$z^* = \min_{c,q,x,f} \sum_{g \in \mathcal{G}} c_g \tag{3.1a}$$

$$\sum_{g \in \mathcal{G}_i} q_{g,t} - D_t^i = \sum_{\substack{l \in \\ from(i)}} f_{l,t} - \sum_{\substack{l \in \\ to(i)}} f_{l,t} \qquad \forall i \in \mathcal{N}, t \in \mathcal{T} \tag{3.1b}$$

$$(c, q, x)_g \in \mathcal{X}_g \qquad \forall g \in \mathcal{G} \tag{3.1c}$$

$$f \in \mathcal{F} \tag{3.1d}$$

The auction model (3.1) aims at minimizing the cost of satisfying the load D_t^i for each time period $t \in \mathcal{T}$ and each bidding zone $i \in \mathcal{N}$. To simplify the exposition of the chapter, demand is assumed to be inelastic[6]. The market includes a set of $\mathcal{G}_i$ suppliers (or market offers) at each node i. Each offer is

5 This has one exception: the so-called PUN orders (the "Prezzo Unico Nazionale" requirement in Italy, cf. NEMO Committee (2020b)) and complex orders are not compatible with the pricing approaches considered in this chapter as they include primal-dual constraints. We point out that both the PUN and complex orders are planned to be discontinued (MCSC, 2023). See also the discussion in section 2.5.

6 All the pricing schemes and results of this chapter can be extended straightforwardly to a model with *elastic* loads, for example a model similar to what is developed in section 2.4. With elastic load, the objective of the auction is welfare maximization (cf. section 2.6).

modelled with a total cost variable c_g, a power output $q_{g,t}$ at time t and a set of possibly non-convex constraints $\mathcal{X}_g$. The variables x_g stand for all the binary variables encountered in the supplier model. In a US auction, which typically relies on a unit commitment model, $\mathcal{X}_g$ should be understood as a detailed representation of the technical constraints of the power plant g. In the EU day-ahead auction, which relies on *portfolio* bidding instead of *unit* bidding, $\mathcal{X}_g$ should be understood as the constraints of the market order g (blocks, linked blocks, stepwise curves, etc.).[7] Equation (3.1b) represents the market clearing constraints. Finally, the auction model (3.1) also includes a network. The variable $f_{l,t}$ represents the flow on line l, while $from(i)$ is the set of lines originating from i and $to(i)$ the ones directed towards i. No assumption is made on the network constraints $\mathcal{F}$, except that it is a *convex* set. All suppliers are assumed to be *price-takers* and to act so as to maximize their private profit. We now proceed with some definitions.

Definition 3.1 (Supplier Profit Maximization)**.** *The agent g is assumed to maximize its selfish profit function $\mathcal{P}_g$, under market price π, defined as follows:*

$$\max_{(c,q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(c,q,x,\pi) \equiv \max_{(c,q,x)_g \in \mathcal{X}_g} \sum_{t \in \mathcal{T}} q_{g,t}\pi_{i(g),t} - c_g. \tag{3.2}$$

Definition 3.2 (Network Profit Maximization)**.** *The network is assumed to maximize its profit function $\mathcal{P}_N$ (the "congestion rent"), under market price π, defined as follows:*

$$\max_{f \in \mathcal{F}} \mathcal{P}_N(f,\pi) \equiv \max_{f \in \mathcal{F}} \sum_{i \in \mathcal{N}, t \in \mathcal{T}} -\pi_{i,t} \left(\sum_{\substack{l \in \\ from(i)}} f_{l,t} - \sum_{\substack{l \in \\ to(i)}} f_{l,t} \right). \tag{3.3}$$

Definition 3.3 (Competitive Walrasian Equilibrium)**.** *The allocation (c^*, q^*, x^*, f^*) together with the market price π constitute a competitive Walrasian equilibrium if*

(i) for each supplier g, $(c^*, q^*, x^*)_g$ optimizes the profit problem (3.2) *under price π ; f^* optimizes the network profit problem* (3.3) *under price π, and*

(ii) the market clears (constraint (3.1b)*).*

A paramount desideratum for an auction is to reach *economic efficiency*: the allocation of goods resulting from the market should be welfare-maximizing (cost-minimizing under inelastic load). All the pricing schemes considered in this chapter assume a welfare-maximizing allocation: they

7 Cf. the developement of section 2.5.

assume that the auctioneer solves problem (3.1) and selects the welfare-maximizing allocation. An example of a pricing scheme that departs from welfare maximization is the current European pricing rule (cf. section 3.1). In the remainder of this chapter, (c^*, q^*, x^*, f^*) refers to the optimal solution of problem (3.1). Since the market is non-convex, a competitive equilibrium is not guaranteed to *exist* (i.e. the concern is about the *existence* of an equilibrium rather than its *efficiency*: the First Theorem of Welfare Economics does not require convexity, so if an equilibrium exists in a non-convex market, it will be efficient, cf. Debreu (1959)). By assumption, the allocation (c^*, q^*, x^*, f^*) satisfies condition (ii) in Definition 3.3. The issue is that there may be no price π that fulfils condition (i), provided this allocation. Assuming that the market agents maximize their profit (Definition 3.1 and 3.2), the violation of condition (i) is measured by the *lost opportunity cost* (LOC).

Definition 3.4 (Lost Opportunity Cost). *The lost opportunity cost is the difference between the maximum profit and the as-cleared profit under price π. It is defined hereafter for each supplier g (eq.* (3.4)*), for the network (eq.* (3.5)*) and in total (eq.* (3.6)*).*

$$LOC_g^{gen}(\pi) = \max_{(c,q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(c,q,x,\pi) - \mathcal{P}_g(c^*,q^*,x^*,\pi) \qquad (3.4)$$

$$LOC^{net}(\pi) = \max_{f \in \mathcal{F}} \mathcal{P}_N(f,\pi) - \mathcal{P}_N(f^*,\pi) \qquad (3.5)$$

$$LOC(\pi) = \sum_{g \in \mathcal{G}} LOC_g^{gen}(\pi) + LOC^{net}(\pi) \qquad (3.6)$$

The lost opportunity cost measures the financial incentives that each profit-maximizing agent has for deviating from the allocation decided by the auctioneer. Having a price that is *incentive compatible* is important to ensure that the participants would follow the dispatch instructions after the market has cleared. Concretely, incentive compatibility is related to the notion of *self-scheduling*: a positive LOC means that the price does not support the dispatch, thereby implying an opportunity for the concerned agents to self-schedule, thus deviating from the dispatch (c^*, q^*, x^*) that is cleared in the auction. As far as the network LOC is concerned, Garcia et al. (2020) interpret it as a potential congestion revenue shortfall, meaning a possible inadequacy between the FTR payments and the congestion revenue that the system operator collects. More generally, it can be interpreted as an incentive for the grid operator, given the market prices, to organise the flows on the network in a manner that deviates from its efficient usage. For example, let us consider two nodes connected by a line. The two nodes receive different prices, but the line is not congested. This could arguably be contemplated as an undesirable configuration (to be contrasted with the ideal situation described in section 2.6.2). Formally, there is a network LOC: the cleared flows do not maximize the value of the network.

Certain researchers and practitioners have advocated that the price should not only aim at being *incentive-compatible*, as measured by the LOC, but that it should also ensure a *non-confiscatory* outcome: the price should at least enable the cleared bids to recover their costs (such non-confiscatory pricing schemes are presented in Madani and Papavasiliou (2022); Bichler et al. (2022); EPRI (2019)). The later is measured by *revenue shortfall*.

Definition 3.5 (Revenue Shortfall)**.** *The revenue shortfall (RS) corresponds to the payments that are required in order to ensure a non-negative profit. It is defined for each supplier (eq.* (3.7)*), for the network (eq.* (3.8)*) and in total (eq.* (3.9)*).*

$$RS_g^{gen}(\pi) = -\min\left(0,\ \mathcal{P}_g(c^*, q^*, x^*, \pi)\right) \tag{3.7}$$

$$RS^{net}(\pi) = -\min\left(0,\ \mathcal{P}_N(f^*, \pi)\right) \tag{3.8}$$

$$RS(\pi) = \sum_{g \in \mathcal{G}} RS_g^{gen}(\pi) + RS^{net}(\pi) \tag{3.9}$$

Needless to say that the LOC and RS are non-negative numbers. Let us notice that lost opportunity cost and revenue shortfall are sometimes referred to, respectively, as "uplift payments" and "make-whole payments" in the literature. However, this terminology is misleading. Because of the absence of a competitive equilibrium, the auctioneer may indeed resort to some sort of out-of-market discriminatory payments that complement the uniform energy price. For example, several US ISOs pay make-whole payments, while ISO-NE pays lost opportunity costs for committed units (EPRI, 2019). Nonetheless, denoting the LOC as "uplift *payment*" suggests that the LOC only matters for the markets that are actually paying them. Instead, the LOC is a crucial indicator (reflecting the opportunities of self-scheduling), independently from the *actual* payments that are paid by a particular auctioneer. Thus, in this chapter, we do not take a stance on what are the side payments that the auctioneer should pay, i.e. whether the auctioneer should pay LOC or only RS ("make-whole payments"), whether off-line resources should be compensated for their lost opportunities, etc. These are important questions, which we nevertheless leave outside the scope of our discussion.[8]

8 Although important, these issues are partly independent of the choice of the uniform market price: there would be no logical contradiction in having an auctioneer using convex hull pricing (see *infra*: the price that minimizes LOC) but who pays RS.

3.3 PRICING SCHEME PROPOSALS

There is no straightforward solution to the absence of competitive prices. We consider hereafter four pricing mechanisms that are proposed in the literature. They all correspond to a certain convex reformulation (either a *relaxation* or a *restriction*) of the non-convex problem (3.1), cf. the discussion in Madani and Papavasiliou (2022). A first option is to rely on *marginal pricing* (O'Neill et al., 2005), also called Integer Programming (IP) pricing. This pricing scheme is theoretically meaningful to study since it is widely used in economics. It is also practically relevant, given its historical usage in US power auctions, and considering that it is a serious candidate currently on the table for the EU market.

Definition 3.6 (Marginal Pricing)**.** *The marginal (IP) prices are the dual variables π^{IP} associated with the market clearing constraint in problem* (3.1) *in which the binary variables x have been fixed to their optimal value x^*.*

It effectively corresponds to taking the price as the subgradient of the total cost curve with binary variables fixed.

A second approach—central for this chapter and for the remainder of the thesis—is Convex Hull Pricing (CHP), which has been proposed in Hogan and Ring (2003) and Gribik et al. (2007). We adopt here the primal formulation of CHP (Hua and Baldick, 2017).

Definition 3.7 (Convex Hull Pricing)**.** *The convex hull prices are the dual variables π^{CH} that are associated to the market clearing constraints in problem* (3.1)*, in which the sets $\mathcal{X}_g$ are replaced by* $\text{conv}(\mathcal{X}_g)$*.*

It is worth noting—besides the peculiar name—the natural interpretation of this pricing approach. The very problem of non-convexities is the inexistence of a competitive equilibrium. The logic of this approach is to compute the prices of the *closest convex economy*, in which a competitive equilibrium exists. Remarkably, although most of the economic theory neglects non-convexities, Starr (1969) and Arrow and Hahn (1971), who studied non-convexities in the theory of general equilibrium, adopted convex hull pricing—albeit they do not use this term. The main property of CHP which has justified its interest in power auctions is that it minimizes the LOC (Gribik et al., 2007): they are the prices that are "as incentive-compatible as possible", i.e. that are as close as possible to a competitive equilibrium.

Proposition 3.1 (CHP)**.** *CH prices minimize the total lost opportunity costs, as defined in* (3.6)*.*

All the proofs are in the appendix 3.A of this chapter. From Lagrangian duality theory, one can observe that the LOC corresponds to the *duality* *gap* between the primal solution z^* and the Lagrangian dual function in which the market-clearing constraint (3.1b) is relaxed. Proposition 3.1 then states that CHP is the price (the Lagrangian multiplier) that minimizes the duality gap.

Convex hull prices are notably difficult to compute (Schiro et al., 2015). Therefore, an *approximation* of CHP, called ELMP, has been proposed and is already implemented by several ISOs, as explained in section 3.1.

Definition 3.8 (Extended Locational Marginal Pricing). *The extended locational marginal prices are the dual variables π^{ELMP} that are associated to the market clearing constraints in problem* (3.1)*, in which the sets $\mathcal{X}_g$ are replaced by $\mathcal{X}_g^{[0,1]}$, i.e. the binary constraints on x are relaxed to* $[0,1]$.

In case $\mathcal{X}_g^{[0,1]} = \text{conv}(\mathcal{X}_g)$, ELMP would correspond to the *exact* CHP approach. This is the main justification for ELMP: it is viewed as a *tractable approximation* of CHP. Nonetheless, even though the above equality, $\mathcal{X}_g^{[0,1]} = \text{conv}(\mathcal{X}_g)$, can be guaranteed in certain simple cases, there are some constraints, such as ramp constraints, for which the equality is not straightforward to obtain, and reaching a *tight* formulation in these cases may require the introduction of a substantial number of valid inequalities (Hua and Baldick, 2017).

In a similar spirit as CHP, which minimizes the LOC, a number of researchers have advocated for a price that minimizes the revenue shortfall. In multiple works, O'Neill has proposed the Average Incremental Cost (AIC) pricing (Chen et al., 2020; O'Neill et al., 2023), which aims at finding a "*zero* make-whole payment price" for the suppliers. However, this is not an achievable target for both the suppliers and the loads if the latter are elastic. Indeed, it cannot be guaranteed that we can find a uniform price that ensures *zero* revenue shortfall for all the market participants in a two-sided auction.

Example 3.1 (Impossibility of Zero RS with Elastic Load). *Let us consider an hourly market with a non-convex supplier producing at maximum 200MW for 50€/MWh, and at minimum 100MW. Let us also consider two convex and* elastic *loads: one is willing to consume 90MW for 10,000€/MWh, the other is willing to consume 20MW for 20€/MWh. Because of the minimum output constraint of the supplier (the non-convexity of the present example), the optimum solution is to produce 100MW and to clear respectively 90 and 10MW of the loads. Any price π would result in either a RS for the loads or for the supplier. Indeed, the non-negative as-cleared profit condition implies $\pi \geq 50$ for the supplier and $\pi \leq 20$ for the load, so the set of prices ensuring* zero *RS is empty*[9]*. We notice*

9 The European SDAC clearing rule achieves *zero* revenue shortfall in a two-sided auction. The difference with Example 3.1 is that the SDAC rule does not fix the optimal dispatch: it allows

that, in this example, CHP, ELMP, MMWP or AIC pricing all result in a market clearing price of 50€/MWh, which implies a RS of 300€ for the second load.

Instead of AIC pricing, we shall consider, as the fourth pricing scheme of this chapter, a method that aims at *minimal* make-whole payments (MMWP), proposed by Madani and Papavasiliou (2022), that works with both elastic and inelastic loads[10]. Two variants of MMWP will later be discussed in section 3.5.

Definition 3.9 (Minimal Make-Whole Payments Pricing)**.** *The minimal make-whole payments prices are the dual variables* π^{MMWP} *associated to the market clearing constraints in the following problem:*

$$\min_{k_g^{gen}, k^f} \sum_{g \in \mathcal{G}} k_g^{gen} c_g^* \tag{3.10a}$$

$$(\pi_{i,t}^{MMWP}) \sum_{g \in \mathcal{G}_i} k_g^{gen} q_{g,t}^* - D_t^i = \tag{3.10b}$$

$$k^f \Big(\sum_{l \in from(i)} f_{l,t}^* - \sum_{l \in to(i)} f_{l,t}^* \Big) \qquad \forall i \in \mathcal{N}, t \in \mathcal{T} \tag{3.10c}$$

$$0 \leq k_g^{gen}, k^f \leq 1 \tag{3.10d}$$

Proposition 3.2 (MMWP)**.** *MMWP prices minimize the total revenue shortfall, as defined in eq.* (3.9)*.*

Given Problem (3.1) assumes inelastic load, MMWP will in fact lead to *zero* revenue shortfall.

We conclude the section with three general remarks. Firstly, among the four pricing approaches, IP, ELMP and MMWP are computationally straightforward to obtain, while CHP is notably more challenging to compute. In this chapter, we calculate it using the Level Algorithm which has demonstrated its ability to compute *exact* CHPs for realistic market sizes (Stevens, 2016; Stevens and Papavasiliou, 2022). This will be the subject matter of chapter 4. Secondly, CHP, IP and MMWP are formulation-*independent*, while ELMP is formulation-*dependent*. Two equivalent formulations of the sets $\mathcal{X}_g$ could result in different ELMPs[11]. Thirdly, we notice that both CHP and ELMP keep primal and dual computations distinct, while IP and

a change in the dispatch, and tolerates a possible loss of social welfare, in order to find a price that ensures zero RS. cf. footnote 3 of this chapter.

10 The original presentation of the method by Madani and Papavasiliou (2022) includes elastic loads. We extend the approach to include a network and inelastic loads.

11 Zhao et al. (2021) have challenged the "formulation-independence" of CHP. However, their usage of the term "formulation" departs from ours. By "formulation", we mean here the textbook definition (Wolsey, 1998): let $\mathcal{X}_g \subseteq \mathbb{R}^n \times \mathbb{Z}^m$, then P_1 and P_2 are two *formulations* of $\mathcal{X}_g$ (e.g. two ways to write ramp constraints) if $\mathcal{X}_g = P_1 \cap (\mathbb{R}^n \times \mathbb{Z}^m) = P_2 \cap (\mathbb{R}^n \times \mathbb{Z}^m)$.

MMWP do not. As highlighted by Schiro et al. (2015), this implies that an off-line unit could set the price under CHP or ELMP. It is, nonetheless, unclear to what extent this is an undesirable feature. For example, the principle of a second-price auction, which is contemplated in economics as a sound manner to clear an auction, is that the first *losing* bid sets the price.

3.4 agents' incentives: distributional analysis

The main property of CHP (Proposition 3.1) informs us on the *total* LOC, which is guaranteed to be lower under CHP than under any alternative price. But it says nothing about how the total LOC is distributed among the market participants. This section studies the main properties that can be established mathematically and observed in the numerical simulations. In general, nothing can be said a priori about how each agent will be affected *individually*, depending on the pricing scheme: although the *total* LOCs are lower under CHP, a supplier *may* have a higher LOC under CHP than under the other prices. Nonetheless, some properties can be established about the split of LOC among the three following categories of market participants: the network, the convex suppliers ($g \in \mathcal{G}^C$) and the non-convex suppliers ($g \in \mathcal{G}^{NC}$, with $\mathcal{G} = \mathcal{G}^C \cup \mathcal{G}^{NC}$). Let us notice that both the European auction and the US markets include a convex network and convex suppliers.

Proposition 3.3 (LOC of Convex Agents in IP)**.** *Under IP pricing, all the convex market participants (the convex suppliers $g \in \mathcal{G}^C$ and the network) have a zero LOC.*

Proposition 3.4 (RS of Convex Agents in IP)**.** *Assuming $0 \in \mathcal{X}_g \; \forall g \in \mathcal{G}^C$ and $0 \in \mathcal{F}$, then both the convex suppliers and the network have a zero revenue shortfall under IP pricing.*

These properties follow from the fact that IP prices reflect the marginal cost of on-line units. Since a *convex* supplier is always on-line and does not bear fixed costs, its LOCs are null under marginal prices. Furthermore, since the primal and the IP pricing problems are coupled so that the flows are equal in both problems, the (convex) network does not bear a LOC. These properties are not shared with the other pricing rules.

Proposition 3.5 (Non-Zero LOC of Convex Agents)**.** *Under CHP, ELMP or MMWP, the convex market participants (both the convex suppliers and the network) may have a positive LOC.*

For the sake of completeness, the following result can also be deduced from Propositions 3.1 and 3.3.

Proposition 3.6 (LOC of Non-Convex Agents)**.** *Under CHP, the total lost opportunity cost of the non-convex suppliers* ($\sum_{g \in \mathcal{G}^{NC}} LOC_g^{gen}(\pi)$) *is lower than under IP prices.*

Intuitively, CHP permits to increase the LOC of the network and the convex generators in order to reduce the total LOC[12]. We shall discuss these Propositions in parallel with the results of the numerical simulations. As announced in the introduction of this chapter, we use two different datasets, each having their merits for the properties we seek to illustrate. The first, later denoted as "FERC dataset", is based on public data (Knueven et al., 2020; Krall et al., 2012). The underlying unit commitment model includes minimum up and down time constraints, ramp constraints (including start-up and shut-down ramps), time-dependant start-up costs, no-load costs, and piecewise linear production costs. The model gathers almost 1000 power units, but has no network. This is a market of realistic size, except for the absence of the network. We conduct our analysis over 11 net-load scenarios of 24 periods each, with hourly time step. The "net-load" is the load net of renewable production, which is given exogenously. The second dataset, later denoted as "CWE dataset", is based on non-public data assembled by our team (Aravena and Papavasiliou, 2016; Stevens and Papavasiliou, 2022). It includes a network of 30 bidding zones and 74 power units. The suppliers are modelled using a simpler unit commitment model than the FERC dataset (essentially simplifying the cost structure). We simulate 12 different load profiles (half of which correspond to 24 periods and the other half correspond to 96 periods). Tables 3.1 and 3.2 report the average results of the FERC and CWE simulations respectively. The detailed results per load scenario are available in appendix 3.C. We will focus on IP, CHP and ELMP, and delay the analysis of MMWP until the next section.

As far as the suppliers are concerned, the FERC data include both a share of convex (14%) and non-convex (86%) suppliers. The CWE data only include non-convex suppliers. We observe that the convex suppliers in the FERC case as well as the network in the CWE case have zero LOC under IP pricing (Proposition 3.3). They also have a null RS (Proposition 3.4). We also observe that CHP outperforms the other prices on the total LOC (Proposition 3.1) as well as on RS, although the latter is not guaranteed by the theory. Tables 3.1 and 3.2 also report the proportion of suppliers impacted by LOC as well as the average LOC carried by these suppliers.

12 Similarly, in case all the suppliers are convex and the network is non-convex, then IP pricing guarantees zero LOC for the suppliers, while CHP transfers some of the LOC from the network to the convex generators, in order to ensure a minimum total LOC (Garcia et al., 2020).

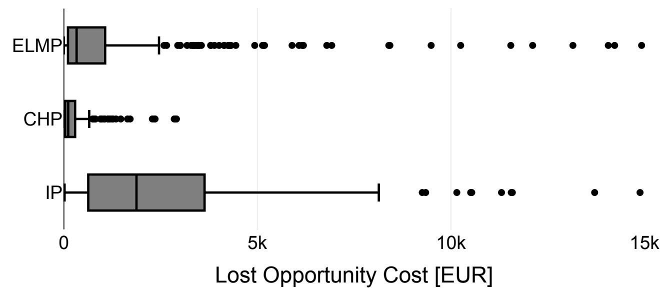


Figure 3.1: Distribution of the LOC across suppliers for IP, CHP and ELMP (aggregate of all the CWE cases).

On the FERC data, we observe that CHP reduces both figures. On the CWE case, the share of suppliers impacted by LOC is similar between CHP and IP pricing, but CHP significantly reduces the the average LOC carried by each supplier (see also Figure 3.1). Interestingly, in both the FERC and CWE datasets, ELMP tends to spread the LOC over a higher share of suppliers.

As far as the network is concerned, we stress two observations. Firstly, Zhao et al. (2021) questions the validity of CHP on the basis that minimizing the network LOC does not make sense. As Propositions 3.3 and 3.5 indicate, it could be argued that CHP minimizes the network LOC to a smaller extent than IP pricing. Secondly, if the concept of network LOC has already been analysed in the literature (Garcia et al., 2020), the concept of network RS has been less discussed. Under some prices, not only could the network bear an LOC (a potential FTR shortfall), but it could also have a shortfall of revenue, i.e. a *negative* congestion rent. The following example illustrates this possibility, although it does not materialize in our simulations. Indeed, Table 3.2 shows that the system operator has positive LOC under CHP, ELMP and MMWP. But the network RS is null under all prices.

Example 3.2 (Network RS)**.** *Let us consider a simple network with two nodes (A and B) connected by a line with a capacity of 100MW. There is an hourly demand of 200MW at 100€/MWh in both nodes as well as a flexible supplier of 400MW at 50€/MWh in node A and an inflexible supplier of 1000MW (all-or-nothing) at 10€/MWh in node B. The welfare-maximizing allocation is to produce 300MWh in node A: 200MWh is consumed in A while 100MWh is consumed in B and the line is congested. Under IP pricing, the prices (π^{IP}) at A and B are 50 and 100€/MWh, respectively and the congestion rent is 5,000€. Under CHP or ELMP, the prices ($\pi^{CHP} = \pi^{ELMP}$) at A and B are 50 and 10€/MWh and the congestion rent is -4,000€.*

		IP	CHP	ELMP	MMWP	MMWP*	MMWP**
Dispatch Cost					29,780,000		
Av. Price [$/MWh]		28.8	28.7	28.8	56.3	26.8	28.9
Num. Suppl. with LOC		3.4%	1.8%	7.5%	79.2%	24.7%	9.5%
Av. LOC per Suppl.		628	19	37	148,232	4,577	94
LOC	Tot.	37,576	323	2,801	130,147,114	1,176,050	14,217
	Conv.	0	67	94	1,978,501	5,268	79
	Non-Conv.	37,576	257	2,707	128,168,613	1,170,782	14,137
RS (in LOC)	Tot.	669	19	206	0	0	0
	Conv.	0	0	3	0	0	0
	Non-Conv.	669	19	203	0	0	0
FO	Tot.	36,907	304	2,596	130,147,114	1,176,050	14,217
	Conv.	0	66	91	1,978,501	5,268	79
	Non-Conv.	36,907	238	2,505	128,168,613	1,170,782	14,137

TABLE 3.1: Incentives of market agents on the FERC dataset depending on the price (average over 11 scenarios). All figures are in US$. Since all the suppliers have the possibility of inaction, $RS_g^{\notin LOC}(\pi) = 0$. The lost opportunity costs (LOC), the revenue shortfall (RS), and the foregone opportunities (FO) are reported for the convex (Conv.) and non-convex (Non-Conv.) suppliers as well as in total (Tot.).

		IP	CHP	ELMP	MMWP	MMWP*	MMWP**
Dispatch Cost					5,489,000		
Av. Price [€/MWh]		42.8	43.4	47.3	27.7	23.8	52.6
Num. Suppl. with LOC		33.2%	35.9%	45.3%	83.6%	63.4%	64.3%
Av. LOC per Suppl.		3,528	278	1,285	141,834	29,326	27,066
LOC	Tot.	83,543	8,093	42,948	98,681,795	41,808,171	20,789,079
	Suppl.	83,543	6,810	39,006	8,746,513	1,350,259	1,250,017
	Net.	0	1,282	3,942	89,935,282	40,457,912	19,539,062
RS (in LOC)	Tot.	10,550	1,987	8,508	0	0	0
	Suppl.	10,550	1,987	8,508	0	0	0
	Net.	0	0	0	0	0	0
FO	Tot.	72,993	6,106	34,440	98,681,795	41,808,171	20,789,079
	Suppl.	72,993	4,823	30,499	8,746,513	1,350,259	1,250,017
	Net.	0	1,282	3,942	89,935,282	40,457,912	19,539,062
RS (not in LOC)	Tot.	897,653	877,040	730,234	0	0	0
	Suppl.	897,653	877,040	730,234	0	0	0
	Net.	0	0	0	0	0	0

TABLE 3.2: Incentives of market agents on the CWE dataset depending on the price (average over 12 scenarios). All figures are in €. The LOC, RS and FO are reported for the suppliers (Suppl.), the network (Net.) and in total (Tot.).

3.5 LOC VS MAKE-WHOLE PAYMENTS CONTROVERSY

As mentioned in section 3.2, some advocate that incentive-compatibility (measured by LOC) is not the adequate target for a price, that should instead aim at being non-confiscatory (measured by RS). Schiro et al. (2015) particularly stress that, in some cases, the revenue shortfall may be lower with IP pricing than with CHP, casting some doubt about the validity of the latter. Although CHP reduces the RS *on average* in our numerical simulations (Tables 3.1 and 3.2), there are indeed instances in both datasets where CHP turns out to modestly increase the RS, cf. appendix 3.C. In order to discuss rigorously the controversy "LOC vs RS", it is first worth clarifying the relationship between LOC and RS.

Proposition 3.7 (Relationship between RS and LOC)**.** *If all the market agents have the possibility of inaction* ($0 \in \mathcal{X}_g \ \forall g \in \mathcal{G}$, $0 \in \mathcal{F}$)*, then* $RS_g^{gen}(\pi) \leq LOC_g^{gen}(\pi) \ \forall g$ *and* $RS^{net}(\pi) \leq LOC^{net}(\pi)$.

Which is to say that, given the possibility of inaction[13], the lost opportunity costs can be viewed as the sum of the revenue shortfall and the foregone opportunities (FO):

$$LOC_g^{gen}(\pi) = RS_g^{gen}(\pi) + FO_g^{gen}(\pi) \quad \forall g \in \mathcal{G}$$
$$LOC^{net}(\pi) = RS^{net}(\pi) + FO^{net}(\pi)$$

The RS is a certain type of LOC in which the cleared profit is negative and the opportunity is to self-schedule at 0, while the FO denotes the remaining "lost opportunities". If the as-cleared profit is zero (as for a unit that is not operating), or positive, the RS is null and the LOC equals the FO, which corresponds to the *additional* profit that the supplier could gain by deviating from the cleared volumes. If the as-cleared profit is negative, the foregone opportunities are the maximal profit *above zero* that the supplier could earn.

Although the possibility of inaction is a standard assumption in economics, there are cases when it does not hold. This happens when there are barriers of exit, for instance, in the presence of must-run constraints (this is the case in Example 7 presented by Schiro et al. (2015)), or in case a supplier that is initially on-line faces a binding "minimum up time" or a ramp constraint that prevents it of being switched off. In these circumstances, Proposition 3.7 does not hold: a unit could produce at a loss ($RS_g^{gen} > 0$) without having any opportunity to act differently ($LOC_g^{gen} = 0$). More specifically, the revenue shortfall could be further dissected into two quantities: $RS_g^{\in LOC}$ (the part of RS which can be expressed as an LOC) and

13 This is the case for all suppliers in the European DA market.

$RS_g^{\notin LOC}$ (the part which cannot be expressed as an LOC, roughly speaking the revenue shortfall due to a barrier of exit). For example, a supplier having an as-cleared profit of −200€ and a maximum profit of 100€, has an LOC of 300€. The latter corresponds to an RS of 200€ as well as an FO of 100€. Alternatively, a supplier which does not have possibility of inaction and which has an as-cleared profit of −200€ and a maximum profit of −100€, has an LOC of 100€ with $RS = RS^{\in LOC} + RS^{\notin LOC} = 100 + 100 = 200$€.

Definition 3.10 (RS & FO)**.** *The revenue shortfall (Definition 3.5) and the foregone opportunities can be further characterised as follows*[14]*:*

$$\begin{aligned} RS_g^{\notin LOC}(\pi) &= \max(0,\ RS_g^{gen}(\pi) - LOC_g^{gen}(\pi)) \\ RS_g^{gen}(\pi) &= RS_g^{\in LOC}(\pi) + RS_g^{\notin LOC}(\pi) \\ FO_g^{gen}(\pi) &= LOC_g^{gen}(\pi) - RS_g^{\in LOC}(\pi) \end{aligned}$$

Under possibility of inaction, $RS_g^{\notin LOC}(\pi) = 0$

CHP minimizes the total lost opportunity costs. Under the possibility of inaction, this means that CHP minimizes the revenue shortfall as long as it does not exacerbate the foregone opportunities. In case the possibility of inaction does not hold, some of the revenue shortfalls ($RS_g^{\notin LOC}(\pi)$) would not enter into what is minimized by CHP. Following those remarks, the LOC-RS controversy, as raised by Schiro et al. (2015), could be formulated as follows:

- Under the possibility of inaction, is it desirable to minimize the RS *at all cost*?
- In case the possibility of inaction does not hold, is it desirable to minimize $RS_g^{\notin LOC}(\pi)$?

We shall present several arguments against both. To address both questions, we rely on the comparison of CHP with MMWP, which is precisely the price that minimizes the RS.

Firstly, is it desirable to minimize the revenue shortfall? A major concern when dealing with MMWP is price indeterminacy: the MMWP prices are typically not unique. This also happens for CHP or IP pricing, as well as for a convex case in which multiple prices could support a competitive equilibrium. Nonetheless, the indeterminacy is expected to be more severe under MMWP than for the other pricing rules. Indeed, minimizing the revenue shortfall is a mild requirement: in a load-inelastic case, *any* price

14 Although we define them for the suppliers, these concepts could also be transposed to the network.

that is high enough would guarantee zero revenue shortfall—e.g. fixing the price at the market price cap would certainly make each cleared bid whole. Mathematically, in problem (3.10), π^{MMWP} belongs to a set that ranges from the smallest price ensuring profitability for all the committed units to infinity.

This indeterminacy is observed in our numerical results. In Table 3.1, the MMWP prices meet their objective of zero revenue shortfall. But this is achieved with prices that are excessively high—two times the CHP on average—which, in turn, leads to extravagant LOC—four times the total system cost. This makes the "vanilla" version of MMWP (Definition 3.9) impracticable. Load elasticity would certainly mitigate the indeterminacy, but it would likely not solve it entirely. If one chooses to proceed with MMWP prices, this then raises the question of how to choose the right price among the *many* MMWP prices. We shall consider two possibilities. The first one, that we shall denote as MMWP*, is to select the *smallest price* that minimizes the revenue shortfall.

Definition 3.11 (MMWP*). *The MMWP* prices are the optimal variables π of the following problem:*

$$\min_{\pi} \|\pi\|_2 \tag{3.11a}$$
$$\mathcal{P}_g(c^*, q^*, x^*, \pi) \geq 0 \quad \forall g \in \mathcal{G} \tag{3.11b}$$
$$\mathcal{P}_N(f^*, \pi) \geq 0 \tag{3.11c}$$

Constraints (3.11b)-(3.11c) require that the price π results in zero revenue shortfall, while the objective (3.11a) resolves the eventual indeterminacy over π by selecting the smallest price that satisfies the required constraints. This method is *akin to* average cost pricing, at least when the load is inelastic, since the smallest price that ensures zero RS is essentially the highest average cost among the committed units. A similar proposal is described by Liberopoulos and Andrianesis (2016).

Bichler et al. (2022) propose another formulation, which we refer to later in the chapter as MMWP**, in which, among the possible MMWP prices, the one that minimizes the LOC is selected. Their model relies on a bi-level optimization problem which is intractable. Consequently, they introduce an approximation of this bi-level model, which consists of finding a price that is as close as possible to ELMP while minimizing the RS[15].

15 The actual model of Bichler et al. (2022) slightly differs from ours: they compute the price that minimizes the RS *for every hour*, as opposed to our model, that minimizes the RS *over the entire market horizon*.

Definition 3.12 (MMWP**)**.** *The MMWP** prices are the optimal variables* π *of problem* (3.11) *in which the objective function (eq.* (3.11a)*) is replaced by* $\|\pi - \pi^{ELMP}\|_2$.

Concretely, MMWP* and MMWP** are linked with MMWP as follows. If $\Pi^{MMWP} = \{\pi \text{ solving } (3.10)\}$, then $\pi^{MMWP*}, \pi^{MMWP**} \in \Pi^{MMWP}$. Finally, we notice that the two previous models are straightforward to extend to a configuration that includes elastic loads, by relying on slack variables in constraints (3.11b)-(3.11c), cf. Bichler et al. (2022).

As far as the numerical results are concerned, we observe in Table 3.1 that, as expected, both MMWP* and MMWP** prices reach zero revenue shortfall. They also both significantly improve the LOC as compared to the vanilla MMWP. Nonetheless, MMWP* is still widely outperformed by the alternative pricing methods. It illustrates that resolving the price indeterminacy that is inherent in MMWP is by no means obvious. This leaves MMWP** as the only serious competitor for IP, CHP and ELMP. We shall nonetheless see later in this section some shortcomings of MMWP** in the CWE case. The question remains: is it desirable to minimize the revenue shortfall *at all cost*? On the FERC simulations, the average total RS under CHP is 19$[16]. Under MMWP**, it drops to zero, but the total LOC increases from 323$ with CHP to 14,217$ with MMWP**. Are the 19$ savings in RS worth the loss of ~14,000$ in LOC? More generally, in the hypothetical case that lowering the RS of 1€ would induce an LOC of 1M€, should we take the stance that minimizes RS? In contrast with MMWP which minimizes the RS *at all cost*, convex hull pricing offers an appealing trade-off: it minimizes the revenue shortfall as long as it does not exacerbate more the foregone opportunities. This is not to say that RS are irrelevant, but since they are unavoidable in two-sided auctions (cf. Example 3.1), considering the above discussion, it may appear more appropriate to handle them through side-payments instead of through the uniform price (as demonstrated by Madani and Papavasiliou (2022), there *always* exist "zero-sum transfers" that can finance the make-whole payments while guaranteeing *revenue-adequacy* for the auctioneer).

Secondly, is it desirable to implement a price that aims at minimizing the RS *including* $RS_g^{\notin LOC}(\pi)$? As a reminder, the three MMWP approaches described so far minimize the total RS, including $RS_g^{\notin LOC}(\pi)$. Let us first look at the question from the viewpoint of a *convex* market. Actually, having $RS_g^{\notin LOC}(\pi) > 0$ is not specific to non-convexities. Indeed, while $LOC = 0$

16 This small number is due to the fact that the FERC dataset does not include network constraints. Including network constraints would likely increase this number, as we observe in the CWE dataset, since the market is "more fragmented", which exacerbates the impact of non-convexities.

is guaranteed in a convex market, it is straightforward to design an instance of a *convex* market (e.g. with a must-run constraint) with a competitive equilibrium, in which some agents have $RS_g^{\notin LOC}(\pi) > 0$. Remarkably, CHP, IP and ELMP would boil down to the classic competitive prices in a convex market, while MMWP would not. Then, the numerical results also highlight another shortcoming of MMWP prices. In the FERC case, all suppliers have the possibility of inaction, and therefore $RS_g^{\notin LOC}(\pi) = 0$. In the CWE case, 36% of the suppliers do not have possibility of inaction because of binding constraints. Consequently, we observe positive $RS_g^{\notin LOC}(\pi)$ in Table 3.2 for all the pricing methods except the three MMWP approaches. We observe that mitigating the $RS_g^{\notin LOC}(\pi)$ through the uniform price of energy comes with a substantial effect on the lost opportunity costs. Intuitively, in order to ensure zero revenue shortfall for suppliers which are in any case not willing to deviate from the market schedule, MMWP raises the prices, which in turn exacerbates the foregone opportunities of the other suppliers. MMWP** which, although disputable, is still competitive in the FERC cases, is simply impracticable in the CWE cases. Again, we are not arguing that the $RS_g^{\notin LOC}(\pi)$ are irrelevant, but according to the evidences of this section, they are not specific to the topic of pricing *non-convexities* and it is not clear that they should be settled through the uniform price of energy, as MMWP does.

3.6 THE LIMITS OF APPROXIMATING CHP

The previous section has focused on MMWP. In the present section, we turn to ELMP. As outlined in section 3.3, the main economic justification for ELMP is that it is viewed as a scalable approximation of CHP which comes with the remarkable Proposition 3.1 (Chao, 2019). This analogy with CHP suggests that ELMP would achieve a lower lost opportunity cost than IP pricing, as it "approximately minimizes LOC". Tables 3.1 and 3.2 confirm this intuition. On average, ELMP roughly cuts by ten (resp. two) the lost opportunity costs in the FERC dataset (resp. CWE dataset) as compared with IP pricing. This is also observed in other works (PJM, 2017; Hua and Baldick, 2017; Yu et al., 2020). Nonetheless, if *empirical* evidence shows that ELMP reduces the LOC as compared to IP pricing, it is worth noting that, in general, there is no *theoretical* guarantee that this will be the case.

Proposition 3.8 (ELMP vs IP LOC). *Given a feasible primal solution of problem* (3.1), *ELMP does not guarantee a lower total LOC than IP pricing.*

Example 3.3 (LOC ELMP vs IP). *Designing a stylized example with* $LOC(\pi^{IP}) < LOC(\pi^{ELMP})$ *is not trivial, since it firstly requires that ELMP differs from CHP.*

Suppliers	x^0	NLC	MC	Q^{max}	Ramp
G1	1	0	80	500	500
G2	0	1950	78	600	300
G3	0	5920	74	600	100
G4	0	0	130	500	105

TABLE 3.3: Supplier data in Example 3.3. The columns stand for the initial commitment, the no-load cost (€/h), the marginal cost (€/MWh), the production limits (MW) and the ramp limits (MW).

D	G1	G2	G4	IP	ELMP	CHP
350	1/350	1/0	0/0	80	80	80
500	1/200	1/300	1/0	80	80	80
950	1/255	1/600	1/95	80	82.5	82.5
1300	1/500	1/600	1/200	180	95.1	145.27

TABLE 3.4: Hourly demand (MW), commitments/schedules (MW) and prices (€/MWh) in Example 3.3.

Let us consider a market with four suppliers (Table 3.3) and four hourly periods with an inelastic load (Table 3.4). The suppliers do not have a minimal production limit, but they have a no-load cost and a ramp constraint (the detailed model is in appendix 3.B). The optimal schedule is reported in Table 3.4. The cheapest way to meet the load in $t = 2$ *is using G1. Nonetheless, due to binding ramp constraints, G2 has to be started in* $t = 1$*, and to produce in* $t = 2$ *in order to meet the ramp from period 2 to 3. Similarly, the cheapest way to satisfy the load in* $t = 3$ *is using G1 and G2. Because of the ramp from period 3 to 4, G4 produces in* $t = 3$*. The total production cost is 267,550€. The binding ramp constraints make ELMP different from CHP. The crux of the example is that G4 has zero no-load cost, as opposed to G3. The optimal schedule commits G4, which has a higher MC, implying a high IP price. In the ELMP pricing problem, since integers are relaxed, the no-load cost of G3 does not have to be borne entirely in periods 2, 3 and 4, rendering it economically more attractive than G4. This drives the ELMP price downward, resulting in a significant revenue shortfall for G4. The prices are reported in Table 3.4 (the intuition about these prices is discussed in appendix 3.B). They lead to a total LOC of 10,670, 12,105 and 3,675€ for IP, ELMP and CHP respectively.*

This is not merely a phenomenon that occurs in a pathological example. In our simulations, there are instances in both datasets where ELMP induces a higher LOC than IP pricing (one instance in both datasets, cf. appendix 3.C). As discussed in section 3.3, ELMP is formulation-dependent. If the formulation of ELMP is tight, then $\pi^{CH} = \pi^{ELMP}$, which implies from Proposition 3.1 that $LOC(\pi^{ELMP}) \leq LOC(\pi^{IP})$. The above discussion highlights that the previous inequality is not guaranteed *in general* for any ELMP, regardless of the tightness of the formulation. This highlights the advantage of *exact* CHP over ELMP, not only for the average reduction of LOC, but also for the theoretical guarantees surrounding CHP. Let us stress that, according to the evidence from Stevens and Papavasiliou (2022), computing *exact* CHP is expected to be feasible for the European market, although this should be confirmed by simulation on the actual order book.

3.7 MINIMIZING THE COSTS OR THE LOC

The last two sections focus on a comparison of IP pricing with CHP, and stress two properties. Again, IP pricing is the candidate currently envisioned by SDAC for the European day-ahead market (MCSC, 2022). Firstly, convex hull pricing minimizes the LOC, not only for the optimal allocation (c^*, q^*, x^*, f^*) of problem (3.1), but for *any* feasible allocation. In this section, we briefly revisit the interplay between primal and dual (pricing) results, also studied in previous works (Sioshansi et al., 2008; Eldridge et al., 2019; Byers and Hug, 2022).

Proposition 3.9 (LOC-Primal Relationship 1). *Under CHP or ELMP, the total LOC decreases monotonically with the optimality gap of the primal solution. More specifically, let $(c, q, x, f)_1$ and $(c, q, x, f)_2$ denote two feasible solutions of problem (3.1), with objectives z_1 and z_2 and lost opportunity cost LOC_1 and LOC_2, respectively. Then:*

$$LOC_1(\pi) - LOC_2(\pi) = z_1 - z_2$$

This result immediately follows from the interpretation of the LOC as the duality gap, explained in section 3.3. Under convex hull pricing, the objective of minimizing the primal optimality gap is consistent with both the minimization of the total costs and the minimization of the lost opportunity costs. Let us notice that Proposition 3.9 holds even if the computation of CHP is not exact. Proposition 3.9 also implies that there is no other allocation that could make the agents better off than the welfare-maximizing allocation. This is notably different under IP or MMWP.

Proposition 3.10 (LOC-Primal Relationship 2)**.** *Under IP or MMWP, the total LOC does not decrease monotonically with the optimality gap of the primal solution.*

Intuitively, as far as IP pricing is concerned, a suboptimal solution commits costlier suppliers which, if entailing higher variable production cost, pulls the IP price upward, which in turn might reduce the LOC. We want to emphasize the dilemma that this might create when it comes to picking the "best" solution among a set of feasible solutions. The dilemma is illustrated on the numerical results of Table 3.5[17]. Here, one instance of the CWE dataset is solved for various optimality gaps. As expected from Proposition 3.9, the LOC associated with CHP and ELMP diminishes monotonically with the primal optimality gap: the improvement in LOC corresponds exactly to the improvement in total cost. Under IP prices, a suboptimal solution (optimality gap of 0.09%) achieves the best LOC. This creates inconsistent incentives for the primal and the pricing problems: going from an optimality gap of 0.09% to 0.01% reduces the total cost by 2,262€ while it increases the lost opportunity cost by 18,139€. Which solution should be preferred? More radically: going from the gap 0.09% to 0.08% reduces the total cost by 826€ while it increases the lost opportunity cost by 93,309€. CHP makes such dilemmas irrelevant.

Another fruitful way of looking at Propositions 3.9 and 3.10 is the following. Economically, in this chapter, we deal with three main requirements: the efficiency of the allocation of resources, the lost-opportunity costs and the revenue shortfalls. We further assume a two-step process in which the auctioneer first selects the optimal allocation of resources (that is, he fully optimizes the efficiency requirement) and he then seeks to find a price that optimizes either the LOC or the RS. One might ask whether there are any losses of generality in this two-step process. For instance, could the LOC be improved by relaxing the efficiency requirement? Propositions 3.9 and 3.10 provide a firm answer to this question: relaxing the efficiency requirement would *not* improve the LOC, under CHP or ELMP. However, this is not true for IP or MMWP pricing, which creates the dilemma discussed above.

3.8 THE CURSE OR BLESSING OF MARKET SIZE

Convex hull pricing does not only minimize the lost opportunity cost, it is also guaranteed to remain *bounded*, so that it does not grow with the market size. This remarkable property, which builds on works from the

17 Since the comparison of this section focuses on CHP and IP pricing, we omit the three MMWP schemes from Table 3.5. Nonetheless, the reader may find the related results for MMWP in appendix 3.A (Table 3.7).

Opt. gap	Tot. Cost	IP LOC	ELMP LOC	CHP LOC
0.1%	5,213,357	115,043	43,346	12,611
0.09%	5,212,947	101,212	42,937	12,201
0.08%	5,212,121	194,521	42,111	11,375
0.07%	5,212,121	194,521	42,111	11,375
0.06%	5,211,690	129,455	41,680	10,944
0.05%	5,211,057	119,929	41,047	10,312
0.04%	5,210,885	119,579	40,875	10,140
0.03%	5,210,743	119,360	40,733	9,997
0.02%	5,210,685	119,351	40,675	9,940
0.01%	5,210,685	119,351	40,675	9,940

TABLE 3.5: Sensitivity of lost opportunity cost to the primal optimality gap, depending on the price. The simulations are performed on CWE dataset (Spring WD 24). All figures are in €.

theory of general equilibrium (Starr, 1969; Arrow and Hahn, 1971), can be expressed for the market model (3.1), in order to derive a theoretical bound on the LOC (Chao, 2019).

Proposition 3.11 (LOC Bound 1)**.** *Under CHP or ELMP, the total LOC is bounded. The bound depends on the shape of $\mathcal{X}_g$, but is independent of $|\mathcal{G}|$:* $\lim_{|\mathcal{G}|\to\infty} LOC(\pi) < \Gamma$.

The surprising feature of Proposition 3.11 is that the LOC does not depend on the market size: if the market grows (increasing the number of suppliers as well as load), given that the LOC remains bounded, its relative importance shrinks ($LOC(\pi^{CH})/z^* \to 0$). The strength of Proposition 3.11 is better captured when contrasted to alternative prices (see also the discussion in Stevens et al. (2024b), reproduced in chapter 5 of this thesis).

Proposition 3.12 (LOC Bound 2)**.** *Under IP or MMWP pricing, the total LOC is not necessarily bounded: it could be that* $\lim_{|\mathcal{G}|\to\infty} LOC(\pi) \to \infty$.

Propositions 3.11 and 3.12 highlight the theoretically sound behaviour of CHP, as opposed to IP pricing. Stylized examples as well as numerical illustrations of these Propositions have nonetheless been scarce in the literature. Example 3.4 aims at providing intuition about the Propositions, while the subsequent numerical simulations and the related discussion explore their practical implications.

Market Size			Convex Hull Pricing		Marginal Pricing	
Number of Plants	Av. Hourly Load (MW)	Tot. Cost ($)	LOC ($)	LOC (% Tot. Cost)	LOC ($)	LOC (% Tot. Cost)
50	4,900	1,820,308	11,222	0.62%	276,383	15.18%
100	9,800	3,631,286	13,114	0.36%	538,713	14.84%
150	14,700	5,444,099	16,841	0.31%	805,370	14.79%
200	19,600	7,245,546	9,202	0.13%	1,060,574	14.64%
250	24,500	9,052,185	6,756	0.07%	1,320,763	14.59%
300	29,400	10,857,007	2,492	0.02%	1,579,297	14.55%
350	34,300	12,666,418	2,817	0.02%	1,842,613	14.55%
400	39,200	14,475,824	3,136	0.02%	2,105,629	14.55%
450	44,100	16,290,191	8,417	0.05%	2,373,870	14.57%
500	49,000	18,099,571	8,711	0.05%	2,636,708	14.57%
1000	98,000	36,183,999	2,280	0.01%	5,258,840	14.53%

TABLE 3.6: Results of CHP and IP pricing on FERC datasets (load profile `2015-08-01_lw`) depending on the market size. The initial 50-unit market is multiplied by a factor ranging from 2 to 20.

Example 3.4 (LOC Bounded or Unbounded). *Consider a session of the European day-ahead market with one hourly period and the following supply orders: one divisible stepwise curve of 100MW at 50€/MWh and a set of N fully indivisible block orders of 100MW at 100€/MWh. Let us assume a divisible demand of 250MW at 1000€/MWh. The welfare maximizing allocation is to clear 2 blocks and 50MW of the stepwise curve. Under IP pricing, the price is 50€/MWh and the two cleared blocks have a revenue shortfall of 10,000€. Let us now assume that the demand grows to 550MW. The IP price remains the same while the revenue shortfall is now 25,000€. This quantity will keep growing with the demand. Under CHP, the price is 100€/MWh. Only the stepwise supply curve has an LOC (in this case, a foregone opportunity) of* $50 \times 50 = 2,500$*€, whether the demand is 250 or 550MW. This shall remain bounded if the demand keeps growing.*

In order to further illustrate the theoretical Propositions, we conduct the following experiment on the FERC dataset over one load profile (`2015-08-01_lw`). First, we randomly select 50 power units out of the 1000. Then, we adapt the load profile accordingly, in order to make the problem feasible. Under these settings, we compute the welfare maximizing allocation as well as the marginal prices and the convex hull prices together with their associated lost opportunity costs. Finally, we gradually increase the market size by duplicating x times the 50 units and multiplying the load accordingly. The results are reported in Table 3.6. We proceed with certain observations. Proposition 3.11 establishes that, when the market size increases, the LOC under convex hull pricing remains bounded and the bound is not affected by the number of plants. Thus, the ratio of the LOC relative to some measure of the market size (e.g. the relative duality gap) is expected to shrink with the market size. This is what we observe in Table 3.6 where the ratio of the LOC relative to the total system cost ranges from 0.62% to 0.01% while the number of power plants grows from 50 to 1000. On the other hand, Proposition 3.12 establishes that the LOC under IP pricing is *not* subject to such a bound and *could* therefore increase with the market size so that the relative importance of LOC remains largely unaffected. Concretely, what we observe in Table 3.6 (last column), is that the ratio of LOC relative to the total system cost remains around 15%, regardless of the market size.

We make two more remarks on the Propositions and the numerical results: the first concerns the mathematical bound in Proposition 3.11, the second concerns the practical implications of the propositions. As far as the bound is concerned, the mathematical expression of Γ is provided in appendix 3.A. This expression can be used to calculate the bound on the FERC dataset: $\Gamma = 21.9$M\$. From Table 3.6, we observe that this bound is far from tight, since the actual LOC amounts to a few thousand dollars per day. Although the trend expected from Propositions 3.11 and 3.12 materializes in the numerical results, the practical usefulness of the bound itself appears to be limited.

As far as the practical implications are concerned, it is of course unrealistic to expect the market to grow by a factor of ten in most US markets or in Europe. We nevertheless stress that the variations of volume traded in the market do not necessarily represent a physical change of generation. In a country such as India, in which the day-ahead market has been created in 2008, and which has recently adopted a similar pricing rule as in Europe (N-SIDE, 2021), such an increase is not far from reality. Indeed, since its creation, the market daily average traded volume has increased by a factor of ten (IEX, 2020). Similarly, in Japan, the traded volume in the day-ahead market was multiplied by more than ten since the implementation of liberalization policies in 2016 (JPEX, 2023). In Europe, if the growth of the day-ahead market is more modest (+1.5% of daily traded volume between 2018 and 2021, with a notable increase of +7% in the number of non-convex block orders over the same period, cf. NEMO Committee (2022)), the traded volume in a market session can vary significantly. As an example, the daily average traded volume in 2020 ranges from 3.83 to 5.82 TWh (NEMO Committee, 2022).

3.9 Conclusion

We have reviewed and analysed six pricing methods from the literature. They are all potential candidates for reforming the current European pricing rule. Marginal pricing could be an upgrade as compared to the current SDAC pricing rule, given the likely improvement in both welfare and scalability. Nonetheless, the fact that many US markets have exhibited the tendency to move away from marginal pricing during the last ten years is something that stakeholders may wish to pay attention to in Europe, given the favourable alternatives that are on the table.

In the chapter, we have attempted to highlight some of the advantages of convex hull pricing over several dimensions. With respect to IP pricing, the fact that CHP incorporates the lumpy costs in the price signal significantly improves the incentives faced by the market agents (section 3.4). CHP is also accompanied by appealing theoretical guarantees, both in terms of consistency between cost and LOC minimization (section 3.7) as well as in terms of the bound on the LOC (section 3.8). While ELMP would be a significant first step in the direction of CHP—a step that several US ISO have made—we have tried to highlight some limits of this approximation. In particular, ELMP does not safeguard all the theoretical guarantees of CHP (section 3.6), nor does it achieve the same performance in terms of LOC minimization. Finally, while minimizing the revenue shortfall—or "make-whole payments"—may appear as a reasonable target, we have shown that it may also result in unbearable (and unbounded, cf. section 3.8) lost opportunity costs (section 3.5).

Throughout this chapter, we have assumed that the market participants would bid truthfully in the auction. This is of course a significant simplification that neglects all the strategic behaviors. A future line of inquiry could be the study of the same problem under strategic settings, leveraging tools from game theory and mechanism design.

APPENDICES

3.A PROOFS OF THE PROPOSITIONS

Proof of Proposition 3.1. Building on Lagrangian duality theory (Wolsey, 1998), CHP as defined in Definition 3.7 is equivalent to solving the following Lagrangian relaxation (Hua and Baldick, 2017).

$$L(\pi) = \min_{\substack{(c,q,x)_g \in \mathcal{X}_g \\ \forall g \in \mathcal{G},\, f \in \mathcal{F}}} \sum_{g \in \mathcal{G}} c_g - \sum_{\substack{i \in \mathcal{N} \\ t \in \mathcal{T}}} \pi_{i,t} \Big(\sum_{g \in \mathcal{G}_i} q_{g,t} - D_t^i - \sum_{l \in from(i)} f_{l,t} + \sum_{l \in to(i)} f_{l,t} \Big) \tag{3.12a}$$

$$\pi^{CH} = \arg\max_{\pi} L(\pi) \tag{3.12b}$$

Hence π^{CH} minimizes the following duality gap:

$$\sum_{g \in \mathcal{G}} c_g^* - \max_{\pi} L(\pi) = \sum_{g \in \mathcal{G}} c_g^* - \max_{\pi} \Bigg[\sum_{i \in \mathcal{N}, t \in \mathcal{T}} \pi_{i,t} D_t^i - \sum_{g \in \mathcal{G}} \max_{(c,q,x)_g \in \mathcal{X}_g} \left\{ \sum_{t \in \mathcal{T}} q_{g,t} \pi_{i(g),t} - c_g \right\} - \max_{f \in \mathcal{F}} \left\{ \sum_{i \in \mathcal{N}, t \in \mathcal{T}} -\pi_{i,t} \left(\sum_{l \in from(i)} f_{l,t} - \sum_{l \in to(i)} f_{l,t} \right) \right\} \Bigg]$$

Replacing D_t^i by $\sum_{g \in \mathcal{G}_i} q_{g,t}^* - \sum_{l \in from(i)} f_{l,t}^* + \sum_{l \in to(i)} f_{l,t}^*$ (using (3.1b)) and rearranging terms, the previous expression is equivalent to

$$\min_{\pi} \left\{ \sum_{g \in \mathcal{G}} LOC^{gen}(\pi) + LOC^{net}(\pi) \right\}.$$

□

Proof of Proposition 3.2. Using a similar result from Lagrangian duality theory as in the CHP approach, computing the prices π^{MMWP} from problem (3.10) is equivalent to solving the Lagrangian relaxation of problem (3.1) in which the sets of constraints are changed from $\mathcal{X}_g$ to $\widehat{\mathcal{X}_g} = \{(0,0,0),\ (c^*,q^*,x^*)_g\}$ and from $\mathcal{F}$ to $\widehat{\mathcal{F}} = \{0,\ f^*\}$. Indeed, the previously defined sets can be modelled with binary variables k. Since solving the Lagrangian relaxation amounts to finding the convex hull of the non-relaxed constraints, and since $\text{conv}(\{0,1\}) = [0,1]$, this leads to problem (3.10). The Lagrangian relaxation is expressed as follows:

$$\min_{\pi} \left\{ \max_{f \in \widehat{\mathcal{F}}} \left\{ \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} -\pi_{i,t} \left(\sum_{l \in from(i)} f_{l,t} - \sum_{l \in to(i)} f_{l,t} \right) \right\} \right.$$
$$\left. + \sum_{g \in \mathcal{G}} \max_{(c,q,x)_g \in \widehat{\mathcal{X}_g}} \left\{ \sum_{t \in \mathcal{T}} \pi_{i(g),t} q_{g,t} - c_g \right\} - \sum_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} \pi_{i,t} D_t^i \right\}$$

Let us replace $D_t^i = \sum_{g \in \mathcal{G}_i} q^*_{g,t} - (\sum_{l \in from(i)} f^*_{l,t} - \sum_{l \in to(i)} f^*_{l,t})$ into the previous expression and let us add the constant $\sum_{g \in \mathcal{G}} c_g^*$. The Lagrangian relaxation then corresponds to:

$$\min_{\pi} \left\{ \sum_{g \in \mathcal{G}} \left(\max_{(c,q,x)_g \in \widehat{\mathcal{X}_g}} \mathcal{P}_g(c,q,x,\pi) - \mathcal{P}_g(c^*,q^*,x^*,\pi) \right) \right.$$
$$\left. + \max_{f \in \widehat{\mathcal{F}}} \mathcal{P}_N(f,\pi) - \mathcal{P}_N(f^*,\pi) \right\}$$

which, from the definition of the modified sets, corresponds to the total revenue shortfall (Definition 3.5). □

Proof of Proposition 3.3. Let us consider the Lagrangian relaxation $L^{IP}(\pi)$ of the problem of Definition 3.6 in which the market clearing constraint is relaxed. Since the problem is convex, the duality gap is zero and $\pi^{IP} = \arg\max_{\pi} L^{IP}(\pi)$. Furthermore, the optimal dispatch of both the primal problem (3.1) (z^*) and the IP problem of Definition 3.6 (z^*_{IP}) is the same: $\sum_{g \in \mathcal{G}} c_g^* = z^* = z^*_{IP}$. We then write:

$$\begin{aligned}
0 &= \sum_{g \in \mathcal{G}} c_g^* - \max_{\pi} L^{IP}(\pi) = \sum_{g \in \mathcal{G}} c_g^* - L^{IP}(\pi^{IP}) \\
&= \sum_{g \in \mathcal{G}^C} \underbrace{\max_{(c,q,x^*)_g \in \mathcal{X}_g} \mathcal{P}_g(c,q,x,\pi^{IP}) - \mathcal{P}_g(c^*,q^*,x^*,\pi^{IP})}_{=LOC_g^{gen} \geq 0} \\
&+ \sum_{g \in \mathcal{G}^{NC}} \underbrace{\max_{(c,q,x^*)_g \in \mathcal{X}_g} \mathcal{P}_g(c,q,x,\pi^{IP}) - \mathcal{P}_g(c^*,q^*,x^*,\pi^{IP})}_{\geq 0, \text{ but } \neq LOC_g^{gen}} \\
&+ \underbrace{\max_{f \in \mathcal{F}} \mathcal{P}_N(f,\pi^{IP}) - \mathcal{P}_N(f^*,\pi^{IP})}_{=LOC^{net} \geq 0}
\end{aligned}$$

From which we conclude that $LOC^{net} = 0$ and $LOC_g^{gen} = 0 \; \forall g \in \mathcal{G}^C$. □

Proof of Proposition 3.4. The result follows from Propositions 3.3 and 3.7. □

Proof of Proposition 3.7. Let us consider the case where $RS_g > 0$ (the unit g faces a revenue shortfall—the case for which $RS_g = 0$ is trivial since $LOC_g \geq 0$):

$$\begin{aligned} LOC_g^{gen} &= \overbrace{\max_{(c,q,x)_g \in \mathcal{X}_g} \left\{ \sum_{t \in \mathcal{T}} q_{g,t} \pi_{i(g),t} - c_g \right\}}^{\geq 0 \text{ by assumption of possibility of inaction}} - \left(\sum_{t \in \mathcal{T}} q^*_{g,t} \pi_{i(g),t} - c^*_g \right) \\ &\geq -\left(\sum_{t \in \mathcal{T}} q^*_{g,t} \pi_{i(g),t} - c^*_g \right) = RS_g^{gen} \end{aligned}$$

The same reasoning applies to the network. □

Proof of Proposition 3.9. This follows the interpretation of the LOC as the duality gap (cf. Proposition 3.1):

$$LOC_1(\pi) - LOC_2(\pi) = \sum_{g \in \mathcal{G}} c_g^1 - L(\pi) - \sum_{g \in \mathcal{G}} c_g^2 + L(\pi) = z_1 - z_2$$

where $L(\pi)$ is the Lagrangian function defined in (3.12a). The equality follows from the fact that CHP and ELMP prices are not affected by a change of primal solution, so the $L(\pi)$ cancel out. □

Proof of Proposition 3.10. The poof for IP pricing derives from the mere observation of Table 3.5. The proof for the three MMWP pricing schemes is straightforward from the observation of Table 3.7, which reports the results of MMWP for the same experience as in Table 3.5. In Table 3.7, we observe that the LOC under MMWP pricing evolves non-monotonically with respect to the primal optimality gap. □

Proof of Proposition 3.11. Ignoring the network, the bound takes the following form:

$$\sum_{g \in \mathcal{G}} LOC_g(\pi^{CH}) \leq \rho |\mathcal{T}|$$

with $\rho = \max_{g \in \mathcal{G}} \rho_g$ and ρ_g defined as follows:

$$\begin{aligned} \rho_g &= \max_{(\hat{c},\hat{q},\hat{x})_g \in \text{conv}(\mathcal{X}_g)} \left\{ \tilde{c}_g(\hat{q},\hat{x}) - \hat{c}_g \right\} \\ \tilde{c}_g(\hat{q},\hat{x}) &= \min_{\substack{(c,q,x)_g \in \mathcal{X}_g \\ q_{g,t} \geq \hat{q}_{g,t}}} c_g \end{aligned}$$

Opt. Gap	MMWP LOC	MMWP* LOC	MMWP** LOC
0.1%	127,174,509	46,374,970	25,487,688
0.09%	128,078,572	46,380,214	25,477,625
0.08%	128,503,837	46,534,306	25,374,010
0.07%	128,503,837	46,534,306	25,374,010
0.06%	129,671,679	46,505,121	25,366,511
0.05%	129,665,324	46,286,384	25,414,900
0.04%	129,855,305	46,366,246	25,411,805
0.03%	127,937,157	46,371,886	25,411,662
0.02%	127,677,227	46,360,290	25,411,605
0.01%	127,677,227	46,360,290	25,411,605

TABLE 3.7: Sensitivity of the LOC under MMWP pricing with respect to the primal optimality gap. The simulations are performed on CWE dataset (Spring WD 24). All figures are in €.

The proof, deriving from an application of the Shapley-Folkman theorem, can be found in Chao (2019) or in Stevens et al. (2024b), cf. appendix 5.A of chapter 5. □

Proof of Proposition 3.12. The proof for IP pricing follows from Example 3.4. A similar stylized example can prove the Proposition for MMWP. Let us consider an hourly market with one fully indivisible block order A of 50MW at 100€/MWh and $N = 3$ block orders B_i of 100MW at 75€/MWh with a minimum acceptance of 90MW. The demand is 240MW at 1,000€/MWh. The welfare maximizing allocation (with or without free disposal) is to clear A as well as two blocks B (one produces 100MW, the other 90MW). In order to ensure zero revenue shortfall, $\pi^{MMWP} = 100$€/MWh. At this price, the blocks B which are not cleared have a foregone opportunity. Clearly, if $N \to \infty$, $LOC(\pi^{MMWP}) \to \infty$. □

3.B MODEL OF EXAMPLE 3.3

The model is the following:

$$\min_{q,x,v,w} \sum_{g\in\mathcal{G},t\in\mathcal{T}} MC_g q_{g,t} + NLC_g x_{g,t}$$

$$\sum_{g\in\mathcal{G}} q_{g,t} = D_t \qquad \forall t \in \mathcal{T}$$

$$0 \leq q_{g,t} \leq Q_g^{max}(x_{g,t} - v_{g,t}) \qquad \forall g \in \mathcal{G}, t \in \mathcal{T}$$

$$q_{g,t+1} \leq q_{g,t} + Ramp_g \qquad \forall g \in \mathcal{G}, t < |\mathcal{T}|$$

$$q_{g,t+1} \geq q_{g,t} - Ramp_g \qquad \forall g \in \mathcal{G}, t < |\mathcal{T}|$$

$$v_{g,t} - w_{g,t} = x_{g,t} - x_{g,t-1} \qquad \forall g \in \mathcal{G}, t > 1$$

$$v_{g,1} - w_{g,1} = x_{g,1} - x_g^0 \qquad \forall g \in \mathcal{G}$$

$$x_{g,t}, v_{g,t}, w_{g,t} \in \{0,1\} \qquad \forall g \in \mathcal{G}, t \in \mathcal{T}$$

where x, v and w stand respectively for the commitment, the start-up and shut-down decision variables. We notice that IP prices (Table 3.4) can be interpreted as follows. G1 is marginal in $t \in \{1,2,3\}$, so π^{IP}=80€/MWh. Increasing the demand of ϵ in $t = 4$ requires increasing the production of G4 in $t = 4$ as well as substituting production of G1 by G4 in $t = 3$, because of the ramp. So $\pi_4^{IP} = 130 + (130 - 80) = 180$€/MWh. ELMP prices (Table 3.4) are less straightforward to interpret, as it is necessary to resort to the KKT conditions of the ELMP problem. To provide some intuition, we look at the price of the third period. The average cost of G2 is $MC + NLC/Q^{max} = 81.25$€/MWh. Increasing the demand of ϵ in $t = 3$ requires to increase the production and commitment of G2 in $t = 3$ as well as to substitute production from G1 by G2 in $t = 2$, so $\pi_3^{ELMP} = 81.25 + (81.25 - 80) = 82.5$€/MWh.

3.C DETAILED NUMERICAL RESULTS

Tables 3.8 and 3.9 provide the detailed results (per load scenario) of Tables 3.1 and 3.2.

		IP	CHP	ELMP	MMWP	MMWP*	MMWP**
2015-02-01-hw	Av. Price	23.1	23.7	23.1	247	20.3	23.2
	Num. Suppl.	7.4%	3.3%	7.2%	99.9%	17.2%	9.1%
	Av. LOC/Suppl.	476	22	53	623,873	8,853	52
	LOC	32,858	673	3,522	582,073,937	1,425,275	4,421
	RS (in LOC)	0	41	88	0	0	0
	FO	32,858	633	3,434	582,073,937	1,425,275	4,421
2015-04-01-hw	Av. Price	19.3	18.9	19.2	27.1	17.3	19.3
	Num. Suppl.	3.4%	1.2%	8.4%	51.7%	16.7%	8.7%
	Av. LOC/Suppl.	265	18	74	41,666	1,546	75
	LOC	8,734	229	6,084	21,082,899	252,047	6,360
	RS (in LOC)	2,426	0	831	0	0	0
	FO	6,307	229	5,253	21,082,899	252,047	6,360
2015-05-01-hw	Av. Price	24.8	24.7	24.8	23.1	23.1	24.8
	Num. Suppl.	1.3%	1.3%	4.1%	60.4%	17.9%	4.1%
	Av. LOC/Suppl.	68	4	12	40,060	1,488	12
	LOC	888	60	471	23,675,307	260,410	471
	RS (in LOC)	499	0	0	0	0	0
	FO	389	60	471	23,675,307	260,410	471
2015-06-01-hw	Av. Price	27.2	27.4	27.1	23.1	23.1	27.2
	Num. Suppl.	2.4%	1.8%	5.3%	64.5%	19.9%	5.2%
	Av. LOC/Suppl.	344	15	20	43,962	6,790	18
	LOC	7,906	271	1,026	27,739,814	1,323,996	923
	RS (in LOC)	0	5	32	0	0	0
	FO	7,906	265	995	27,739,814	1,323,996	923
2015-07-01-lw	Av. Price	32.8	32.9	32.8	49.6	32.8	32.9
	Num. Suppl.	1.2%	1.1%	4.4%	83.7%	31.6%	5.1%
	Av. LOC/Suppl.	231	21	29	38,983	4,543	35
	LOC	2,772	241	1,273	31,926,833	1,403,848	1,733
	RS (in LOC)	21	0	5	0	0	0
	FO	2,751	241	1,268	31,926,833	1,403,848	1,733

2015-07-01-hw	Av. Price	28.6	27.8	28.7	26.4	27.1	28.7
	Num. Suppl.	2.4%	3.3%	8.8%	60.4%	23.8%	8.8%
	Av. LOC/Suppl.	608	13	42	36,043	2,611	42
	LOC	13,978	427	3,583	21,301,239	608,479	3,583
	RS (in LOC)	3,922	0	0	0	0	0
	FO	10,056	427	3,583	21,301,239	608,479	3,583
2015-08-01-hw	Av. Price	28	27.2	28.1	23.1	26.4	28.1
	Num. Suppl.	3.2%	1.4%	11.0%	81.3%	31.8%	11.1%
	Av. LOC/Suppl.	749	24	31	75,557	2,156	30
	LOC	23,217	336	3,341	60,067,559	670,588	3,327
	RS (in LOC)	229	12	38	0	0	0
	FO	22,988	324	3,303	60,067,559	670,588	3,327
2015-09-01-lw	Av. Price	43.3	43	43.4	34.3	41	43.6
	Num. Suppl.	4.1%	2.7%	9.6%	83.1%	44.7%	14.9%
	Av. LOC/Suppl.	168	18	26	49,988	7,778	99
	LOC	6,719	468	2,461	40,640,061	3,399,179	14,509
	RS (in LOC)	233	0	82	0	0	0
	FO	6,486	468	2,379	40,640,061	3,399,179	14,509
2015-09-01-hw	Av. Price	35.2	36.9	35.3	78.6	33.2	35.8
	Num. Suppl.	8.8%	1.3%	12.7%	99.2%	33.7%	21.9%
	Av. LOC/Suppl.	3,579	29	35	241,607	5,830	511
	LOC	307,764	383	4,318	234,358,740	1,923,877	109,366
	RS (in LOC)	0	71	435	0	0	0
	FO	307,764	313	3,883	234,358,740	1,923,877	109,366
2015-10-01-lw	Av. Price	30	30.3	30	61	27.4	30.2
	Num. Suppl.	2.4%	1.4%	4.6%	90.8%	22.1%	6.4%
	Av. LOC/Suppl.	366	26	46	105,984	7,301	57
	LOC	8,053	341	1,973	89,874,328	1,503,907	3,403
	RS (in LOC)	0	0	91	0	0	0
	FO	8,053	341	1,882	89,874,328	1,503,907	3,403
2015-12-01-hw	Av. Price	23.8	23.8	23.8	26.1	23.2	23.9
	Num. Suppl.	1.0%	1.0%	6.9%	96.1%	12.2%	9.0%
	Av. LOC/Suppl.	50	14	43	332,826	1,447	99
	LOC	447	128	2,763	298,877,534	164,941	8,286
	RS (in LOC)	26	85	660	0	0	0

FO	421	43	2,104	298,877,534	164,941	8,286

TABLE 3.8: Incentives of market agents on the FERC dataset depending on the pricing scheme (detailed figures per scenario).

		IP	CHP	ELMP	MMWP	MMWP*	MMWP**
SpringWE-24	Av. Price	35.6	36.4	43.8	25.3	24.2	50.3
	Num. Suppl.	28.4%	25.7%	37.8%	78.4%	60.8%	60.8%
	Av. LOC/Suppl.	4,047	278	2,631	137,534	36,071	38,786
	LOC	84,978	7,189	75,852	95,284,255	45,775,853	27,790,203
	RS (in LOC)	1,207	1,145	15,035	0	0	0
	FO	83,771	6,043	60,816	95,284,255	45,775,853	27,790,203
	RS (not in LOC)	1,241,106	1,262,137	937,572	0	0	0
AutumnWE-24	Av. Price	38	38.6	45.2	30.8	24.4	51.1
	Num. Suppl.	29.7%	23.0%	36.5%	78.4%	55.4%	56.8%
	Av. LOC/Suppl.	4,245	195	2,164	228,052	36,306	36,682
	LOC	93,398	4,814	61,033	134,946,417	45,300,887	25,886,284
	RS (in LOC)	9,812	1,364	10,898	0	0	0
	FO	83,586	3,450	50,135	134,946,417	45,300,887	25,886,284
	RS (not in LOC)	1,165,408	1,177,396	880,050	0	0	0
SummerWE-24	Av. Price	34.5	34.5	42.4	25.1	23.9	49.6
	Num. Suppl.	29.7%	23.0%	37.8%	78.4%	63.5%	63.5%
	Av. LOC/Suppl.	5,549	621	2,861	115,019	36,695	40,142
	LOC	122,078	12,606	82,506	88,196,688	46,897,168	29,419,777
	RS (in LOC)	6,111	4,985	23,074	0	0	0
	FO	115,967	7,621	59,431	88,196,688	46,897,168	29,419,777
	RS (not in LOC)	1,231,897	1,312,780	997,636	0	0	0
SummerWE-96	Av. Price	44.3	44.4	46.8	24.9	21.1	51.4
	Num. Suppl.	41.9%	36.5%	52.7%	81.1%	70.3%	71.6%
	Av. LOC/Suppl.	2,286	212	627	99,952	22,600	17,941
	LOC	70,879	6,406	25,444	76,307,676	35,056,506	15,473,856
	RS (in LOC)	3,814	2,858	7,065	0	0	0
	FO	67,065	3,547	18,379	76,307,676	35,056,506	15,473,856
	RS (not in LOC)	634,481	654,700	577,425	0	0	0

SummerWD-24	Av. Price	35.3	34.2	42.9	23.2	24.2	49.9
	Num. Suppl.	25.7%	25.7%	39.2%	79.7%	60.8%	63.5%
	Av. LOC/Suppl.	3,612	357	2,450	106,374	37,784	39,207
	LOC	68,620	7,707	73,911	86,363,710	46,951,981	29,096,219
	RS (in LOC)	7,190	3,832	21,504	0	0	0
	FO	61,430	3,875	52,408	86,363,710	46,951,981	29,096,219
	RS (not in LOC)	1,319,226	1,288,743	984,371	0	0	0
AutumnWD-24	Av. Price	47.9	43.4	49.6	33.6	26.7	55
	Num. Suppl.	32.4%	40.5%	41.9%	81.1%	58.1%	56.8%
	Av. LOC/Suppl.	4,839	464	817	198,139	32,992	33,969
	LOC	116,130	18,723	42,626	134,179,378	45,844,078	23,812,415
	RS (in LOC)	67,061	649	1,025	0	0	0
	FO	49,070	18,074	41,601	134,179,378	45,844,078	23,812,415
	RS (not in LOC)	1,048,066	888,203	832,867	0	0	0
AutumnWD-96	Av. Price	53	52.6	54	30.1	25.8	58.1
	Num. Suppl.	35.1%	55.4%	58.1%	94.6%	70.3%	71.6%
	Av. LOC/Suppl.	1,999	152	388	126,419	20,139	14,407
	LOC	51,962	6,625	21,179	90,412,716	39,833,338	13,737,122
	RS (in LOC)	7,886	1,194	1,146	0	0	0
	FO	44,076	5,430	20,033	90,412,716	39,833,338	13,737,122
	RS (not in LOC)	538,906	532,300	507,761	0	0	0
SpringWE-96	Av. Price	44.8	45.3	47.3	25.7	21.1	51.7
	Num. Suppl.	37.8%	44.6%	50.0%	86.5%	67.6%	68.9%
	Av. LOC/Suppl.	2,841	131	626	198,060	22,316	17,651
	LOC	79,560	5,196	24,688	94,436,434	35,076,443	15,125,326
	RS (in LOC)	7,511	633	4,893	0	0	0
	FO	72,049	4,563	19,795	94,436,434	35,076,443	15,125,326
	RS (not in LOC)	627,309	640,835	561,709	0	0	0
SummerWD-96	Av. Price	46.7	46.1	49.1	25.7	23.2	53.5
	Num. Suppl.	36.5%	43.2%	55.4%	83.8%	71.6%	73.0%
	Av. LOC/Suppl.	2,238	165	556	90,557	22,047	15,174
	LOC	60,430	5,989	24,358	80,534,617	38,674,968	14,967,433
	RS (in LOC)	3,356	1,339	4,493	0	0	0
	FO	57,074	4,650	19,865	80,534,617	38,674,968	14,967,433
	RS (not in LOC)	645,930	629,168	556,458	0	0	0

SpringWD-24	Av. Price	44.5	42.9	47.4	30.4	25	53.3
	Num. Suppl.	28.4%	32.4%	37.8%	82.4%	52.7%	54.1%
	Av. LOC/Suppl.	5,683	400	1,185	182,800	37,165	36,427
	LOC	119,351	9,940	40,675	127,677,227	46,360,290	25,411,605
	RS (in LOC)	6,463	3,829	8,973	0	0	0
	FO	112,889	6,111	31,702	127,677,227	46,360,290	25,411,605
	RS (not in LOC)	1,110,721	969,030	859,078	0	0	0
AutumnWE-96	Av. Price	46	45.2	48.3	25.4	21.6	52.6
	Num. Suppl.	39.2%	39.2%	45.9%	83.8%	63.5%	64.9%
	Av. LOC/Suppl.	2,581	167	643	88,274	25,366	18,805
	LOC	74,843	5,591	23,722	82,089,816	36,154,282	14,545,232
	RS (in LOC)	5,296	1,116	2,911	0	0	0
	FO	69,546	4,475	20,811	82,089,816	36,154,282	14,545,232
	RS (not in LOC)	625,250	614,743	543,490	0	0	0
SpringWD-96	Av. Price	50	49.6	51	32.2	24.2	55.1
	Num. Suppl.	33.8%	41.9%	50.0%	94.6%	66.2%	66.2%
	Av. LOC/Suppl.	2,412	189	467	130,828	22,438	15,598
	LOC	60,292	6,328	19,382	93,752,601	39,772,253	14,203,481
	RS (in LOC)	895	898	1,079	0	0	0
	FO	59,397	5,430	18,303	93,752,601	39,772,253	14,203,481
	RS (not in LOC)	583,540	554,445	524,396	0	0	0

TABLE 3.9: Incentives of market agents on the CWE dataset depending on the pricing scheme (detailed figures per scenario).

4

COMPUTATIONAL METHODS FOR CONVEX HULL PRICING

ABSTRACT. *Convex hull pricing is a well-documented method for coping with the non-existence of uniform clearing prices in electricity markets with non-convex costs and constraints. We revisit primal and dual methods for computing convex hull prices, and discuss the positioning of existing approximation methods in this taxonomy. We propose a dual decomposition algorithm known as the Level Method and we adapt the basic algorithm to the specificities of convex hull pricing. We benchmark its performance against a column generation algorithm that has recently been proposed in the literature. We provide empirical evidence about the favorable performance of our algorithm on large test instances based on PJM and Central Europe markets.** [*]

KEYWORDS. Convex hull pricing · Non-uniform pricing · Level method · Bundle methods

JEL CLASSIFICATION. C61 · C63 · C68 · D47 · Q41

* The chapter reproduces, with minor changes, the content of Stevens and Papavasiliou (2022). The most significant changes include a reworking of the introduction, changes in the usage of some concepts and notations to improve the consistency with chapter 3, as well as the addition of an appendix, section 4.A, which provides the reader with an illustration of the algorithms discussed in the chapter on a 2-D example.

4.1 INTRODUCTION

The classical analysis of an economic dispatch problem, together with its dual, provides a fundamental argument for *uniform pricing* in electricity markets (cf. section 2.6): an optimal dispatch can be supported by a set of competitive equilibrium prices. In other words, even if a central authority cannot effectively control the dispatch of the assets itself, it can provide prices that align the behaviour of selfish profit maximizing agents with social welfare maximization. However, as the argument assumes convexity of the dispatch problem, a *fundamental* challenge is *non-convexity*, as the latter implies that it is not guaranteed that a competitive market equilibrium exists.

Non-convexities are at the heart of power system operations, in terms of both the *network model* as well as in the *market orders* (Taylor, 2015): (i) they are present in the alternating current (AC) power flow equations which characterize the physics of the grid and (ii) in the mixed integer programming (MIP) constraints that describe the market offers. As the day-ahead (DA) markets in Europe and in the US rely on a linear direct current (DC) power flow model of the grid, point (i) is not encountered in these markets[1]. On the other hand, point (ii) is a reality in both US markets that rely on solving a unit commitment (UC) problem, as well as in the EU market which includes integer market orders, such as the so-called "block orders"[2]. Throughout this chapter, we neglect (i) and rather focus on (ii).

The inexistence of equilibrium prices in electricity auctions has triggered a long-lasting debate on the choice of an appropriate pricing scheme in the presence of non-convexities. Convex hull pricing (CHP) has arisen as one promising alternative: while being so far mainly debated in the US, it has also recently emerged as a possible option for the EU market (NEMO Committee, 2020a). Chapter 3 has undertaken a broad economic analysis of the different pricing proposals that have been made in the literature. This analysis has highlighted several advantages of convex hull pricing. In this chapter, we turn to the computational aspects of convex hull pricing. Indeed, a practical concern of CHP is that its computation can be challenging, as often ackowledged in the literature (e.g. Issue 7 in Schiro et al. (2015)). This chapter aims at addressing these computational challenges by putting forward a workable algorithm—the *Level Method*—for realistic instances subject to network constraints. In the remainder of this section, we repeat the main background concepts related to CHP, covered

1 Note, nevertheless, that the debate on TSO/ISO-DSO integration has recently motivated the consideration of more advanced models for the representation of network constraints in market-clearing platforms (Garcia et al., 2020).

2 Cf. section 2.5.

in chapter 3, as well as the context of non-uniform pricing discussions in the EU. Insofar as the EU market is concerned, we focus the discussion on computational issues, which motivate our choice of test instances.

NON-UNIFORM PRICING SCHEMES. The most widely debated "non-uniform pricing schemes" in the literature, reviewed in chapter 3, include integer programming (IP) pricing (O'Neill et al., 2005), convex hull pricing (Hogan and Ring, 2003; Gribik et al., 2007), and "extended LMP" (ELMP) pricing which has been applied early on in the PJM market (PJM, 2017; Federal Energy Regulatory Commission, 2019). They all amount to a *convex reformulation* of the market clearing problem. These strategies consist of combining a *uniform electricity price* with discriminatory payments, called *uplift payments*, which aim at restoring the incentives of market participants for following the market matches. In this framework, the overall market clearing procedure can be described in three steps, which are also followed by our simulations:

1. Solve the *primal problem*, in order to establish the dispatch and commitment instructions ;

2. Solve a *pricing problem* in order to compute uniform electricity prices ;

3. Solve the independent *profit maximization problems* of all market agents (generators and the network operator) in order to establish *uplift payments*.

Regarding step 3, although the definition of uplift payments is controversial and varies across ISOs[3], we shall focus in this chapter on the lost opportunity costs (LOC). The main justification for this choice is computational rather than economical. Indeed, this chapter studies the *computation* of convex hull pricing, which minimizes the LOC. Therefore, the LOC turns out to be a convenient indicator for measuring the convergence of the algorithms computing the convex hull prices. As discussed in chapter 3, the lost opportunity costs can be split between network LOC, and suppliers' LOC, each defined as the difference between the maximum profit achieved by self-scheduling given the market prices and the as-cleared profit. The total LOC is the sum of these two quantities[4].

Regarding the above step 2, IP pricing is a common choice in non-convex settings (cf. the historical account covered in chapter 3). We also use it as a benchmark for our simulations. However, it does not attempt to minimize lost opportunity costs, and can therefore possibly lead to high

3 Cf. chapter 3.
4 Cf. Definition 3.4.

side payments. The lost opportunity costs are undesirable, as they can distort the incentives of bidders or create revenue adequacy problems for the market operator that needs to finance them (Van Vyve, 2011). These concerns—together with other appealing properties advocated in chapter 3—motivate *Convex Hull Pricing* (CHP), the main property of which is to minimize lost opportunity costs[5]. Because it is computationally challenging, PJM (and other US ISOs) has recently implemented a new pricing scheme, referred to as "extended LMP" which is more tractable computationally than CHP. For certain forms of simple market orders, it can also be shown to be a reasonable *approximation* of CHP (PJM, 2017). We expand on how it relates to the computation of CHP in section 4.2.

UNIFORM PRICING IN THE EU. The EU market landscape presents a number of major institutional differences compared to US markets[6]. One such notable difference is that day-ahead energy auctions are operated by *for-profit Nominated Electricity Market Operators* (NEMOs) while, in the US, it is the (typically *non-profit*) ISO that operates both the market and the network. One implication of this difference relates to the ability of the market operator to socialize uplift payments. This difference may, in part, justify the currently employed "uniform" pricing scheme that is adopted in Europe, as implemented in Euphemia, the algorithm that clears the pan-European day-ahead auction (NEMO Committee, 2020b).

In Euphemia parlance, the aforementioned generator lost opportunity costs can be related to: (i) *paradoxically accepted blocks* (PAB)—cleared bids actually facing losses, i.e. requiring *make-whole payments* (cf. section 3.2)—and (ii) *paradoxically rejected blocks* (PRB)—a rejected bid that would have been profitable, i.e. facing a *foregone opportunity*. The EU day-ahead market "avoids" uplift payments by (i) constraining the problem by not allowing the acceptance of PABs while (ii) allowing PRBs, but not paying their lost opportunity costs. Ultimately, it does not effectively reduce the lost opportunity costs to zero, but it guarantees zero *make-whole payments*, while increasing the total *lost opportunity cost* and not paying it—which creates self-scheduling opportunities for the market participants. Consequently, this pricing scheme only outputs *uniform prices* while it does not provide the market participants with any discriminatory payments. This justifies why, in EU NEMO parlance, it is referred to as *uniform*, in contrast to the three *non-uniform* pricing schemes that are discussed previously.

This *uniform* pricing scheme involves "primal-dual" constraints that implicate dispatch and price decisions in a single market clearing model.

5 cf. Proposition 3.1
6 cf. the discussion in section 3.1

The solution implemented in Euphemia amounts to an iterative algorithm that matches market orders while aiming to find a feasible price (without PAB). If this is not possible, the algorithm generates a cut in the primal model and repeats the process. In contrast to the *non-uniform* pricing schemes that work in three steps (dispatch, price, uplifts), the EU "uniform" pricing scheme works as a single—but *iterative*—step, and couples dispatch and price problems together.

This makes the problem that Euphemia is called to solve (a mixed integer quadratic program subject to complementarity constraints) computationally challenging. Moreover, the approach deteriorates market welfare, since welfare-enhancing orders can be discarded if no market clearing price can be found to support the aforementioned clearing rule. For these reasons, non-uniform pricing schemes, and in particular *convex hull pricing*, have recently received consideration by the European NEMOs as a possible option for the European DA energy auction (NEMO Committee, 2020a). Considering the aforementioned institutional EU structure, as well as the algorithm implemented in Euphemia, this would constitute a disruptive market design evolution.

Computationally speaking, implementing CHP in Europe comes with three paramount requirements (NEMO Committee, 2020b,a):

- Euphemia is afforded 12 minutes of run time[7].
- The market model includes a network of ~ 40 bidding zones, and its geographic footprint is expected to be further enlarged.
- The market model is expected to move towards 15-minute granularity in the near future (a horizon of 96 periods).

Forty bidding zones for ninety-six periods implies a 3,840-dimensional price space. These requirements motivate the considered use cases in section 4.4.

CONTRIBUTIONS AND STRUCTURE OF THE CHAPTER. The contribution of the chapter is twofold:

1. We propose the *Level Method* (Nesterov, 2004) for computing CHP and adapting it to the specificities of our problem. We specifically adapt the algorithm in order to exploit the convexity of the *network* model. We further introduce a "multi-cut" variant of the Level Method in order to leverage the separability of the sub-problems. Note that

7 This held true when the corresponding paper of this chapter was written. As noted in chapter 3, the runtime limit is now 17 minutes. This change neither affects the discussion nor the conclusions of this chapter.

two types of approaches have been envisioned in the literature for solving CHP: *dual approaches* and the *primal approaches* (we define these in section 4.2). The *Level Method* belongs to the former. Primal approaches, and their drawbacks which motivate our choice for a dual approach, are presented in section 4.2. The review of alternative (tested) dual approaches comes in section 4.3 and motivates our choice of the Level Method.

2. We efficiently solve CHP, using the Level Method, for large instances *including a network* and a horizon of 96 periods, which anticipates the evolution of the EU market. We conduct a critical comparison of our approach against both *primal* and *dual decomposition* approaches. In particular, we compare it to a notable recent publication by Andrianesis et al. (2021), which proposes a Dantzig-Wolfe (D-W) algorithm for computing CHP. The D-W algorithm exhibits favorable performance on a test case without a network and with 24 time periods, as considered in Andrianesis et al. (2021). Given our preoccupation with a market clearing model at the scale of the EU market, the question becomes how the method scales when moving from a 24-dimensional to a 3,840-dimensional price space. When increasing the dimension, the Level Method is empirically shown to attain favorable performance relative to Andrianesis et al. (2021).

Our chapter is inspired by an older unpublished work (Stevens, 2016), and is further motivated by Andrianesis et al. (2021). We describe the mathematical formulation of CHP in section 4.2. We then introduce the Level Method in section 4.3. In section 4.4, we test the algorithm on multiple large instances and compare the results with D-W. Section 4.5 concludes and discusses areas of future research.

4.2 MATHEMATICAL FORMULATION

CONVEX HULL PRICING PROGRAM. We define the dispatch problem subject to *network constraints* as follows:

$$\min_{c,q,x,f} \sum_{g \in \mathcal{G}} c_g \tag{4.1a}$$

$$(\pi_t^i) \quad \sum_{g \in \mathcal{G}_i} q_{g,t} - D_t^i = \sum_{\substack{l \in \\ from(i)}} f_{l,t} - \sum_{\substack{l \in \\ to(i)}} f_{l,t} \quad \forall i, t \tag{4.1b}$$

$$(c_g, q_{g,t}, x_{g,t}) \in \mathcal{X}_g \quad \forall g \in \mathcal{G} \tag{4.1c}$$

$$f \in \mathcal{F} \tag{4.1d}$$

Here, $\mathcal{G}_i$ denotes the set of generators (or market offers) at node i. Each offer is modelled with a total cost c_g, a power output $q_{g,t}$ at time t and a set of non-convex constraints $\mathcal{X}_g$. The generic variables x_g stand for all the binary variables encountered in the generator model. The demand at time t and node i, D_t^i, appears in the market clearing (MC) constraints (4.1b). Regarding the network, $f_{l,t}$ stands for the flow on line l, while $from(i)$ is the set of lines originating from i and $to(i)$ the ones directed towards i. No assumption is made on the network constraints $\mathcal{F}$, except that it is a *convex* set.

Each generator g is assumed to be a selfish agent that maximizes profit, i.e. solves the following program:

$$\max_{c,q,x} \sum_t q_{g,t} \pi_t^{i(g)} - c_g \tag{4.2a}$$

$$(c_g, q_{g,t}, x_{g,t}) \in \mathcal{X}_g \tag{4.2b}$$

Here, $i(g)$ stands for the node of generator g, while $\pi_t^{i(g)}$ represents the market price of node $i(g)$ at time t.

A fundamental result on CHP establishes that minimizing lost opportunity costs amounts to solving the following problem (Hogan and Ring, 2003; Gribik et al., 2007):

$$\pi^{CHP} = \arg\max_{\pi} L(\pi) \tag{4.3}$$

Here, $L(\pi)$ denotes the *Lagrangian dual function*, obtained by relaxing constraints (4.1b) of problem (4.1):

$$L(\pi) = \sum_{i,t} \pi_t^i D_t^i \tag{4.4a}$$

$$- \sum_{g \in \mathcal{G}} \max_{(c,q,x)_g \in \mathcal{X}_g} \left\{ \sum_t q_{g,t} \pi_t^{i(g)} - c_g \right\} \tag{4.4b}$$

$$+ \min_{f \in \mathcal{F}} \left\{ \sum_{i,t} \pi_t^i \left(\sum_{\substack{l \in \\ from(i)}} f_{l,t} - \sum_{l \in to(i)} f_{l,t} \right) \right\} \tag{4.4c}$$

We recognize in (4.4b) the profit maximization problems (4.2) of the generators. As established in Gribik et al. (2007), using the optimal primal dispatch solution of (4.1) and injecting it into (4.4) clarifies why the previous Lagrangian problem does indeed minimize the LOC[8]. As also pointed out

8 Cf. the develomements of chapter 3 and in particular Proposition 3.1.

in the literature, the definition (4.3) of CHP also indicates that the LOC can be interpreted as the *duality gap* between (4.1) and (4.4).

US VERSUS EU MODELS. In addition to institutional differences between US and EU markets, another major difference relates to the definition of market products. The US markets follow a *unit-bidding* model, where each unit is represented explicitly in the market, along with its technical characteristics. On the other hand, the EU day-ahead market follows a *portfolio-bidding* model (which cannot be subsumed in the unit commitment formulation), where each agent submits multiple generic market orders that represent the portfolio of its assets in an aggregated way. These market orders include convex hourly orders—*stepwise* and *interpolated curves*—as well as non-convex orders—mainly the family of *block orders*[9] (NEMO Committee, 2020b). The latter is a financial order spanning over multiple periods and involving a *binary acceptance* variable.

Model (4.1) remains general regarding the bid (generator) constraints (4.1c), which are simply represented as the non-convex set $\mathcal{X}_g$. This implies that the approach outlined in this chapter can accommodate all the flavours of unit commitment models as well as the EU-like auctions. This exceeds what a "primal CHP approach" can model.

Finally, model (4.1) considers a general (but *convex*) set of network constraints $\mathcal{F}$. Our approach can in fact accommodate any convex representation of the network. In both the US and EU market, $\mathcal{F}$ would amount to a set of *linear* constraints, the main difference being that certain US markets are *nodal* (larger number of nodes) while the EU market is *zonal* (roughly one zone per country). We remark in section 4.3 on the specific treatment of the network in our proposed Level Method.

PRIMAL AND DUAL APPROACHES FOR COMPUTING CHP. In this section, we locate the Level Method in the perspective of the landscape of all the alternative of approaches for solving CHP and we motivate the choice of a dual approach in light of the limitations of the primal approaches.

As noted in section 4.1, there are two main approaches envisioned for computing convex hull prices—i.e. solving problem (4.3): (i) the *Lagrangian dual approaches*, which directly attempt to maximize function $L(\pi)$ using an iterative algorithm, and (ii) the *primal approaches*, understood as methods that seek to describe the CH of the non-relaxed constraints (4.1c)-(4.1d) by developing tight formulations. Figure 4.1 outlines the landscape of

9 Note that other non-convex (and less standard) products in Euphemia such as the Italian unique national price (PUN) or complex orders NEMO Committee (2020b), are not directly compatible with CHP, because they implicate primal and dual variables in their definition. Cf. section 2.5 for a discussion.

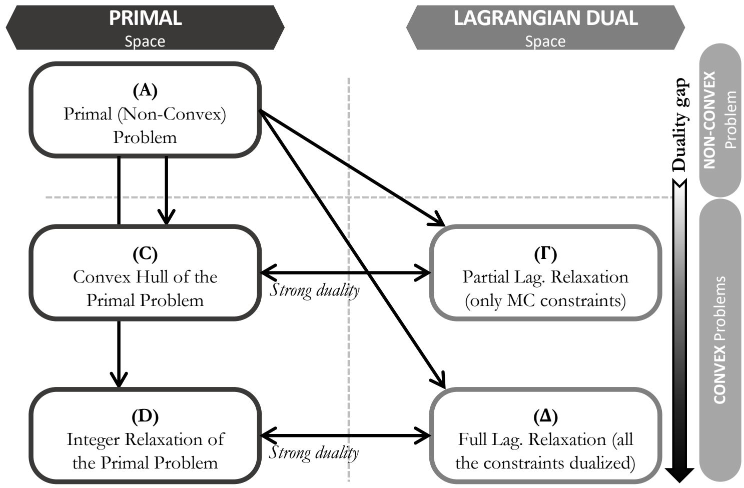


FIGURE 4.1: Landscape of problems for computing or estimating convex hull prices.

approaches for computing convex hull prices. The top problem (A) corresponds to the dispatch problem (4.1). Below, on the left, we find primal relaxations of (A) while, on the right, we find Lagrangian relaxations of (A)—Lagrangians are indeed a widely employed method for deriving convex relaxations of non-convex programs (Lemaréchal, 2001). The problem (Γ) corresponds to the CHP definition (4.3), which can be solved by dual decomposition approaches such as the Level Method. Problem (Γ) maps to its primal equivalent in (C). The underlying idea of the primal formulation is that computing the CHP as the Lagrangian multipliers of (4.3) is equivalent to computing the dual variable π associated to the market clearing constraint (4.1b) in the primal problem (4.1), if the latter is expressed on the convex hull of its domain—i.e. $\text{conv}(\mathcal{X}_g)\ \forall g \in \mathcal{G}$ (see Wolsey (1998) and Lemaréchal (2001) for the general result in Lagrangian relaxation theory, and Hua and Baldick (2017) for the specific result related to CHP).

Although (C) is the tightest primal relaxation of (A), there exist looser relaxations, such as (D), which amounts to relaxing the integrality constraints $x_{g,t} \in \{0,1\}$ to $x_{g,t} \in [0,1]$. This corresponds to *ELMP pricing*, discussed in the introduction. ELMP pricing can be interpreted as a computationally efficient *approximation* of CHP[10]. In certain cases, relaxing the integrality constraints in $\mathcal{X}_g$ may provide $\text{conv}(\mathcal{X}_g)$. In this case, problems (C) and

10 See PJM (2017) as well as the discussion of chapter 3.

(D) are equivalent and ELMP pricing effectively corresponds to convex hull pricing. The fact that relaxation (D) is looser than (C) implies that the *duality gap* between (A)–(D) will be greater than or equal to the one between (A)–(C).

Interestingly, one can also relate the primal relaxed problem (D) to its Lagrangian dual counterpart (Δ). While CHP is solving the *partial* Lagrangian dual relaxation (Γ), ELMP pricing corresponds to solving the *full* (looser) Lagrangian dual relaxation (Δ), where all the constraints—and not only the market clearing constraints—are dualized[11].

Regarding the primal CHP problem (C), a way to approach it is to develop a *tight*—but *custom*—formulation, specific to the targeted problem (A). Recent researches have embraced this idea: Hua and Baldick (2017) proposes an explicit formulation for the primal model of CHP for classical UC constraints. Madani et al. (2018) analyses primal CHP formulations for the constraints of the European day-ahead market clearing model[12]. More recent research further elaborates on the idea, developing tight (custom) formulations for MISO (Yu et al., 2020) or proposing a network flow model of unit commitment, in order to compute CHP for a broader set of constraints (Álvarez et al., 2019). One value of the primal CHP approaches is to establish the link between CHP theory and the literature dedicated to tight formulations of UC polytopes[13]. Similarly, when including a non-convex network model, the primal CHP approach also establishes the connection between CHP theory and SDP–SOCP relaxations of AC power flow (Garcia et al., 2020).

Nevertheless, as also voiced in Andrianesis et al. (2021), there are certain constraints for which the convex hull is not tractable in the sense that it may not be possible to characterize the convex hull with a scalable number of constraints. This already holds for simple ramp constraints (Hua and Baldick, 2017). This is also acknowledged by Álvarez et al. (2019), where the authors do not account for these ramp constraints in their network

11 Taylor (2015) proposes an interesting interpretation of CHP by relating it to the semi-definite programming (SDP) relaxation of problem (4.1). The proposition is motivated by the well-known SDP relaxation of a non-convex quadratically constrained program (QCP) (Vandenberghe and Boyd, 1996; Boyd and Vandenberghe, 1997; Lemaréchal and Oustry, 2001) and the fact that a MIP can be expressed as a QCP. However, the above taxonomy reveals an inaccuracy in the reasoning: it mixes (Δ) and (Γ), as it omits the fact that CHP relies on a partial (and not complete) Lagrangian relaxation, where only the market clearing constraints are relaxed (i.e. dualizing fewer constraints can only improve the duality gap (Lemaréchal and Renaud, 2001)).

12 Note however that Madani et al. (2018) focuses on a subset of the market constraints, ignoring e.g. linked blocks and exclusive groups NEMO Committee (2020b).

13 See (Morales-España et al., 2013, 2015; Gentile et al., 2017; Rajan and Takriti, 2005; Damcı-Kurt et al., 2013; Queyranne and Wolsey, 2017; Silbernagl et al., 2015; Knueven et al., 2020; Sridhar et al., 2013). See also chapter 2 in Stevens (2016) for a critical discussion of this literature.

flow model. Instead, Yu et al. (2020) needs to combine the proposed tight formulation with an iterative algorithm in order to account for the ramp constraints in a *scalable* way. It goes without saying that these modelling limitations also hold for more advanced constraints such as multimode CCGT units, detailed battery models, and so on. Thus, since the pricing mechanism becomes dependent on the quality of the primal formulation, the primal approach can be ruined by adding a new constraint—which is particularly concerning, since electricity market models are constantly subject to changes. These modelling limitations imply that, if the representation of the convex hull is not tractable, the primal approaches are irremediably left with an *approximation* of convex hull prices, such as the ELMP pricing model (D). This is illustrated in our numerical results of section 4.4, where the primal method benchmark of Hua and Baldick (2017) is included. This motivates our choice for a dual approach.

4.3 THE LEVEL METHOD

4.3.1 *Review of existing algorithms*

The appropriate algorithmic scheme for solving (4.3) is related to the type of function $L(\pi)$.

Proposition 4.1 (Concave). *Function $L(\pi)$ is concave in π.*

Proposition 4.2 (Non-smooth). *Function $L(\pi)$ is a non-smooth (piecewise linear) function, where each facet can be seen as corresponding to a set of binary (commitment) decisions x_g.*

Proposition 4.3 (First-order oracle). *A first-order oracle is available, i.e. given a price π, both the function value $L(\pi)$ as well as its supergradient $s \in \hat{\partial} L(\pi)$ can be evaluated.*

Proposition 4.4 (Supergradient). *Let (c^*, q^*, x^*, f^*) be the optimal reactions to π (solving respectively* (4.4b) *and* (4.4c)*). Then*

$$s = D_t^i - \sum_{g \in \mathcal{G}} q_{g,t}^* + \sum_{l \in from(i)} f_{l,t}^* - \sum_{l \in to(i)} f_{l,t}^*$$

is a supergradient of L in π; i.e. $s \in \hat{\partial} L(\pi)$.

Each call to the oracle implies solving MIP profit maximization programs (4.4b) for each generator as well as for the network program (4.4c)—these are thus *slave problems*. We propose later a special treatment of the network and leverage the separability of the profit maximization problems in order to substantially improve the formulation. Any algorithm tackling this problem would work in three steps:

1. Given a price π_k, evaluate $L(\pi_k)$ and $\hat{\partial} L(\pi_k)$;
2. Given this information, generate a new price π_{k+1} ;
3. If the stopping criterion is met, stop. Else, go to step 1.

The main difference between dual decomposition algorithms is in the way that they construct the sequence of iterates $\{\pi_k\}_{k=0}^{\infty}$: (i) some algorithms simply update the prices based on the latest supergradient information—they are *memoryless*—; (ii) while other algorithms will keep memory of the sequence of iterates. We briefly summarize three approaches, which were tested and compared to the Level Method by Stevens (2016).

A well-known scheme belonging to category (i) is the *subgradient scheme*. Perhaps surprisingly, it is proven to be optimal for general convex non-smooth optimization with arbitrarily high dimension (Nesterov, 2004). However, when dealing with problems of "moderate" dimension such as the one presented in our context, there exist more optimistic alternatives.

Indeed, the subgradient scheme for piecewise-linear functions, such as our problem (4.3), tends to oscillate between the facets of the Lagrangian dual function, around an edge. Therefore, one idea is to "catch the edge" and follow it until the optimum, instead of oscillating from one facet to another, as the subgradient method does. This intuitive reasoning leads to the *Extreme-Point Subdifferential* (EPSD) algorithm, which has been specifically applied to the CHP problem (Wang et al., 2013a,b). However, our experiments in Stevens (2016) reveal that each iterate of the algorithm is costly, as it requires not only to solve the problems (4.2) for each generator to optimality, but to enumerate *all* the optimal solutions[14].

Unlike these two *memoryless* schemes, the *Analytic Center Cutting Plane Method*[15] (ACCPM) is based on the principle of iteratively reducing the search domain: the price domain is initially limited to a box and, at each iterate, the supergradient is used for generating a *cut*, which shrinks the search domain. The next testing point is chosen as the analytical center of the updated domain.

Our original investigation of these alternative *dual approaches* (subgradient, EPSD and ACCPM) concluded that none of them were competitive with the *Level Method* for computing CHPs (Stevens, 2016).

---

14 Some illustrations of these algorithms are provided in the appendix 4.A of this chapter.

15 See Nesterov (2004) and Boyd et al. (2008) for the theory and Wang et al. (2013b, 2009) for its application to CHPs.

4.3.2 *Kelley's approach*

The Kelley algorithm forms the basis for the proposed Level Method (Nesterov, 2004). It is based on the idea of iteratively constructing a *model* (an upper approximation) of the Lagrangian function $L(\pi)$, using its supergradients.

Definition 4.1 (Model Function). *Let Q be the initial domain of our problem (i.e. a* box *limiting the prices, which can be economically interpreted as* price caps*) and let $\{\pi_k\}_{k=0}^{\infty}$ be a sequence in Q. Let s_k be the supergradient at iterate π_k. Then*

$$\widehat{L}(\pi, k) = \min_{j=0..k} \{\langle s_j, \pi - \pi_j \rangle + L(\pi_j)\} \tag{4.5}$$

is a model for the Lagrangian function $L(\pi)$, such that $\widehat{L}(\pi, k) \geq L(\pi)$.

In order words, the piecewise linear function $L(\pi)$ is upper-approximated at each iterate by a model function $\widehat{L}(\pi, k)$ consisting of supporting hyperplanes. At iteration 0, this is a single hyperplane. Then, as the iterate count k is increasing, the model function $\widehat{L}(\pi, k)$ is becoming increasingly accurate.

Definition 4.2 (Master Program). *The maximization of the model function yields the master program at iterate k:*

$$\begin{aligned} \max_{\pi \in Q,\ \theta} \quad & \theta \\ s.t. \quad & \theta \leq \langle s_j, \pi \rangle + b_j \quad \forall j = 0..k \end{aligned} \tag{4.6}$$

Here, s_j are the "cut coefficients" (as defined in Property 4.4) and $b_j = L(\pi_j) - \langle s_j, \pi_j \rangle$ are the "cut constants". This is a computationally tractable linear program.

Having the upper-approximation function $\widehat{L}(\pi, k)$ at hand, one needs to decide the rule for building the sequence of iterates $\{\pi_k\}_{k=0}^{\infty}$. The more intuitive way to pick the next iterate is:

$$\pi_{k+1} = \arg\max_{\pi} \widehat{L}(\pi, k). \tag{4.7}$$

i.e. the solution of the master program (4.6). This defines Kelley's cutting plane method. One of its benefits is that it explicitly provides an *upper bound* as well as a *lower bound* at each iterate k: a lower bound is defined as $LB_k = \max_{j=0..k} L(\pi_j)$, while an upper bound is $UB_k = \max_{\pi} \widehat{L}(\pi, k)$. Note that the sequence of upper bounds $\{UB_j\}_{j=0}^{k}$ is decreasing, as the definition of the model function implies that $\widehat{L}(\pi, k+1) \leq \widehat{L}(\pi, k)$. The upper and lower bounds can be combined to define the *relative gap*, which is used as a *stopping criterion* for the Kelley (and Level) Method:

$$\frac{UB_k - LB_k}{|UB_k|} \leq \epsilon \tag{4.8}$$

4.3.3 *Level stabilization*

Kelley's algorithm is *finite*, because each iterate adds a new hyperplane and the number of hyperplanes supporting the function is finite. Nevertheless, despite its simplicity and its good behaviour in low dimension, it tends to be unstable in higher dimension[16]. This is due to the unstable nature of piecewise linear functions: adding a new supporting hyperplane can move the optimum far from the previous point (i.e. to a corner of the box Q). This well-known drawback justifies why multiple *stabilization approaches* have been proposed in the literature, including the *Level Method* (Nesterov, 2004; Frangioni, 2020).

The underlying idea of the Level Method is to update prices more smoothly: instead of using the optimum of the model function as the next iterate, the algorithm chooses π_{k+1} such that it is "better" than π_k, as evaluated by the model function $\widehat{L}(\pi_{k+1}, k)$, without being optimal at all costs. We observe in section 4.4 that this stabilization has a major influence on the practical performance of the algorithm.

A graphical illustration in 1-D is presented in Figure 4.2. The cuts, the LB and the UB are obtained as in Kelley's method, by solving the master program (4.6). However, unlike in Kelley's method, the next price candidate is selected by solving a projection program.

Definition 4.3 (Projection Program)**.** *The iterate π_{k+1} is chosen as the projection of π_k on the "level set" $\widehat{L}(\pi, k) \geq \alpha UB_k + (1-\alpha)LB_k$, which amounts to solving:*

$$\begin{aligned} \min_{\pi \in Q} \quad & ||\pi - \pi_k||_2^2 \\ s.t. \quad & \langle s_j, \pi \rangle + b_j \geq \alpha UB_k + (1-\alpha)LB_k \quad \forall j = 0..k \end{aligned} \tag{4.9}$$

Here, $\alpha \in [0,1]$ is the projection parameter. This is a computationally tractable quadratic program.

Regarding the calibration of α, $\alpha = 1$ corresponds to the classic Kelley method, while $\alpha = 0$ implies that the iterate simply does not move. We note that a theoretically optimal α exists for *general* convex non-smooth functions (Nesterov, 2004), but that a calibration to the *specific* problem can

16 This is illustrated in appendix 4.A.

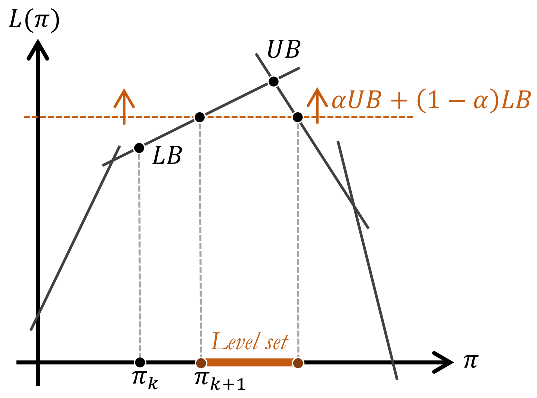


FIGURE 4.2: Illustration of projection on the level set.

still be meaningful. Our empirical tests on the CHP problem reveal that, for the high-dimensional instances that we are interested in, the approach is largely insensitive to the choice of α. This is shown later in Table 4.3, where any value of α between 0.2 and 0.7 exhibits similar performances. Following Stevens (2016), the value $\alpha = 0.2$ is chosen for all of our experiments in the present work.

Regarding the choice of the box Q, experimental evidences show that the Level Method is not too sensitive to its exact value, although it impacts the quality of the UB estimate. In all of our experiments, Q is initially set to $\pm 300\$/MWh$ and is then progressively shrunk after 10, 20 and 30 iterates to $\pm 25\$/MWh$ around the latest price candidate. This is justified by an analysis of the volatility of the price iterates, which rapidly reach a price close to the CHP.

4.3.4 *Refinements of the Level Method in the context of CHP*

We now propose adjustments to the basic algorithm which exploit the structure of our problem. We specifically leverage the fact that: (i) the network model is *convex* and (ii) the profit maximization programs of the generators are *separable*.

In our development so far, we have been treating the *convex* network term (4.4c) identically to the *non-convex* generators, i.e. by solving the network profit maximization given a price π, and generating a supergradient. We illustrate below the treatment of the *convex parts* of the primal program by focusing our discussion on the network. The idea applies identically to *con-*

vex generators (e.g. the convex orders in Euphemia, which are numerous), a *convex* pumped-storage model, etc.[17]

For the sake of illustration, let us assume that the network constraints $\mathcal{F}$ correspond to the *DC (voltage angle) power flow*. Term (4.4c) then reads as follows:

$$\min_{f,\psi} \sum_{i,t} \pi_t^i \left(\sum_{l \in from(i)} f_{l,t} - \sum_{l \in to(i)} f_{l,t} \right) \tag{4.10a}$$

$$(\mu_{l,t}) \quad f_{l,t} \leq \overline{F}_l \qquad \forall l, t \tag{4.10b}$$

$$(\nu_{l,t}) \quad f_{l,t} \geq \underline{F}_l \qquad \forall l, t \tag{4.10c}$$

$$(\lambda_{l,t}) \quad f_{l,t} = B_l(\psi_{or(l),t} - \psi_{dest(l),t}) \qquad \forall l, t \tag{4.10d}$$

Here, B_l stands for the susceptance of line l, and $\overline{F}_l$ and $\underline{F}_l$ are its max and min capacity, while $or(l)$ and $dest(l)$ denote the origin and destination nodes of line l. The dual of (4.10) can be expressed as:

$$\max_{\mu \geq 0, \nu \geq 0, \lambda} \sum_{l,t} \nu_{l,t} \underline{F}_l - \mu_{l,t} \overline{F}_l \tag{4.11a}$$

$$\pi_t^{or(l)} - \pi_t^{dest(l)} + \mu_{l,t} - \nu_{l,t} + \lambda_{l,t} = 0 \qquad \forall l, t \tag{4.11b}$$

$$\sum_{l \in to(i)} \lambda_{l,t} B_l - \sum_{l \in from(i)} \lambda_{l,t} B_l = 0 \qquad \forall i, t \tag{4.11c}$$

Problem (4.11) can now be injected into (4.4) as a substitute for (4.4c), meaning that the network dual variables (μ, ν, λ) would explicitly be variables of the master (and projection) program. This process allows to provide more information directly into the master program, hence improving the available model $\widehat{L}(\pi, k)$ of function $L(\pi)$.

Secondly, the classical Kelley and Level Methods add a *single cut* at each iterate, namely one single cut for all the generators. Nevertheless, the dual function is *separable* with respect to the generators. We therefore propose a *multi-cut Level Method*, whereby we compute one cut (one lower approximation) for each generator profit maximization subproblem. Our experiments reveal that this adaptation can deliver substantial computational benefits. Generating more cuts makes the *model function* more accurate, which enables the algorithm to converge faster. Note that multi-cut versions of other approaches have been applied successfully in different contexts, such as for two-stage stochastic programs (Birge and Louveaux, 1988, 2011).

17 See section 3.6 and appendix A in Stevens (2016) for a treatment of these cases.

To summarise, after the inclusion of both the network dual and the multi-cut approach, the master program (4.6) at iterate k becomes:

$$\max_{\substack{\mu \geq 0, \nu \geq 0, \\ \lambda, \pi \in Q, \theta}} \sum_{i,t} \pi_t^i D_t^i + \sum_{l,t} \left(\nu_{l,t} \underline{F}_l - \mu_{l,t} \overline{F}_l \right) - \sum_{g \in \mathcal{G}} \theta_g \tag{4.12a}$$

$$\theta_g \geq \langle q_{g,\cdot}^j, \pi^{i(g)} \rangle - c_g^j \quad \forall g, j = 0..k \tag{4.12b}$$

$$\pi_t^{or(l)} - \pi_t^{dest(l)} + \mu_{l,t} - \nu_{l,t} + \lambda_{l,t} = 0 \quad \forall l, t \tag{4.12c}$$

$$\sum_{l \in to(i)} \lambda_{l,t} B_l - \sum_{l \in from(i)} \lambda_{l,t} B_l = 0 \quad \forall i, t \tag{4.12d}$$

Here, $\{q_g^j\}_{j=0}^k$, corresponds to the sequence of generator g power output for iterates $j = 0..k$. These parameters are also *cut coefficients* for generator g. On the other hand, $\{c_g^j\}_{j=0}^k$, which corresponds to the sequence of generator g cost for iterates $j = 0..k$, are the *cut constants*. The translation of the projection program (4.9) is applied as discussed previously.

In the classical Kelley or Level Methods, estimating the lower bound (evaluating (4.4) at a given π) follows directly from the resolution of the slave subproblems. The inclusion of the network into the master program, as described above, complicates the process. Indeed, the network contribution in the dual function (4.4c) is not solved explicitly anymore, but now comes in the master objective (4.12a), together with constraints (4.12c) and (4.12d) that should not be violated. Therefore, estimating the value of $L(\pi)$ after having retrieved the cuts from the slaves (for the same π) amounts to solving the master (linear) program (4.12) with the variables π fixed. The overall procedure is described schematically in Figure 4.3. Note that the resolution of the two master programs (with π fixed and variable) can be parallelized.

4.4 simulation results

This section presents the numerical results of the (multi-cut) Level Method on instances of realistic scale. The Level Method has been benchmarked against other *dual approaches* in earlier work by the authors (Stevens, 2016). It is chosen as the most promising method for computing CHPs among all tested alternatives. In the present section, we therefore focus on its comparison with a recent work by Andrianesis et al. (2021), which employs a D-W column generation algorithm (i.e. the dual of Kelley, cf. Vanderbeck and Wolsey (2010)) for iteratively building the convex hull of the dispatch problem, i.e. D-W gradually discovers the corners of the primal formulation. As in the case of the Level Method, it can be applied to *any* UC

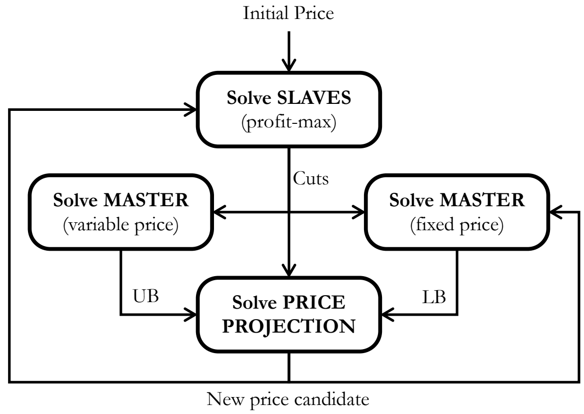


Figure 4.3: Our implementation of the Level Method for the computation of CHP.

formulation. We use it as a performance benchmark in our analysis, due to its favourable empirical performance. We also include IP pricing, discussed in the introduction, as another benchmark in our analysis, as well as ELMP pricing (discussed in section 4.2) as a *primal method* benchmark.

Unlike other computational researches on CHP which are mainly concerned about the number of generators in the problem (Andrianesis et al., 2021; Wang et al., 2013b), we rather focus our investigations on the sensitivity of the algorithms with respect to the dimension of the price space. Indeed, although the number of generators is surely relevant, since the ultimate goal is to compute *prices* by optimizing $L(\pi)$, the price-space dimension is expected to have a significant impact on the performance of any tested method. Therefore, we first present results *without* a network, with a horizon of 24 periods, and then introduce network constraints and extend the time horizon to 96 periods.

For all our test cases, the comprehensive market procedure for computing the prices and measuring lost opportunity costs follows the steps that are described in section 4.1. Concretely, there are three steps: dispatch, price, and lost opportunity costs computation. The Level Method and D-W differ with respect to the second step. Both approaches have been implemented in Julia (JuMP) and all the tests are run on a personal computer (Intel Core i5, 2.6 GHz with 8 GB of RAM) using Gurobi 9.1.1.

Economic results		Computational results	
Dispatch Cost [$]	29,791,214	Level iter	19
IP LOC [$]	652,263	Level av. time/iter[a][s]	8.2 (0.36)
Primal LOC [$]	11,400	D-W iter	29
CHP LOC [$]	9,746	D-W av. time/iter[a][s]	8.9 (0.34)

[a] $(\cdot)$ denotes the average time per iterate for solving the "master programs" (i.e. master plus projection in the case of the Level Method).

TABLE 4.1: Results of the Level Method and the Dantzig-Wolfe algorithm on FERC datasets (average over 11 instances).

4.4.1 *FERC (US) test cases*

The first test cases in our analysis are based on FERC datasets (Knueven et al., 2020; Krall et al., 2012)[18]. The test sets are publicly available, together with the associated UC model, and are also used by Andrianesis et al. (2021), which permits a sound comparison. These test cases consist of a detailed UC model. The only adaptations in our work are the removal of reserve and the netting out of renewable supply from the load. The UC model includes, among others, min up and down time constraints, ramp constraints (including start-up and shut-down ramp rates), variable start-up costs which depend on how long a unit has been off, no-load costs, and piecewise linear production costs. The model has *no network*, but gathers > 930 generators. This corresponds to an instance of realistic size, barring for the absence of the network. As in Andrianesis et al. (2021), we conduct our analysis on a 24-period horizon with hourly time step.

Table 4.1 presents the average results over 11 FERC instances, while Figure 4.4 illustrates the convergence behaviour of both approaches on one of the instances. The 11 instances essentially correspond to 11 different load profiles, with slight changes in the production fleet, which varies from 934 to 978 generators. The stopping criterion of the Level Method (equation (4.8)) is set to 0.01%. The number of iterates reported in Table 4.1 for D-W corresponds to the iterations that are required for reaching the same amount of LOC as the Level Method. Both algorithms are initialized at a uniform price of 20$/MWh.

The results already show the attractive performance of the Level Method, both (i) in terms of *iteration count* and (ii) in terms of *robustness*. Indeed, there is an average improvement of 34% compared to D-W in terms of

18 These are the same as the so-called "FERC dataset" used in chapter 3.

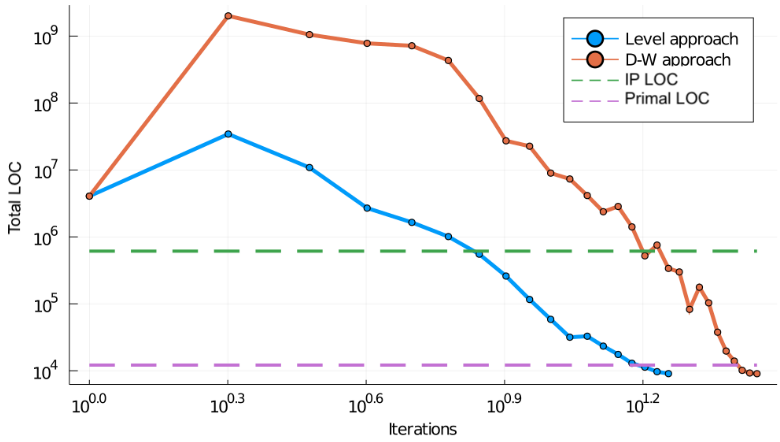


FIGURE 4.4: Convergence of the Level Method and the Dantzig-Wolfe algorithm, measured by the lost opportunity costs (IP pricing and the primal method are used as benchmark thresholds), on the `FERC 2015-07-01 high wind` instance. Both axes are in logarithmic scale.

number of iterates (Table 4.1). It should be noted that this number of iterates is a reasonable measure for comparing the performance of both approaches. Concretely, both methods have to solve the same subproblems and mainly differ in the other computations that they are required to perform. Whereas the Level Method has to solve both a linear master and a quadratic projection, D-W is only required to solve the linear (master) extended formulation. On the other hand, the extended formulation solved by D-W is larger than the Level master program, as illustrated in Figure 4.5. Overall, this results in a similar run time per iterate, as reported in Table 4.1 which shows both the average run time per iterate as well as, between parentheses, the average run time spent in the master programs (master plus projection programs for the Level Method). This implies that the number of iterations (the *analytical complexity*: the number of calls of the oracle to reach a reliability target) is a reasonable measure for comparing performance. It also has the benefit of being less dependent on the specific machine or on the implementation details. Note that, for both approaches, the slave subproblems can be parallelized.

Furthermore, there is a gain in *robustness*: the Level Method exhibits a more stable performance, as observed in Figure 4.4. Indeed, Figure 4.4 suggests that it does not seem possible to stop the D-W algorithm long before its termination, since LOC remains high for a large number of iterations (we also refer the reader to Figure 4.7 of the next use case,

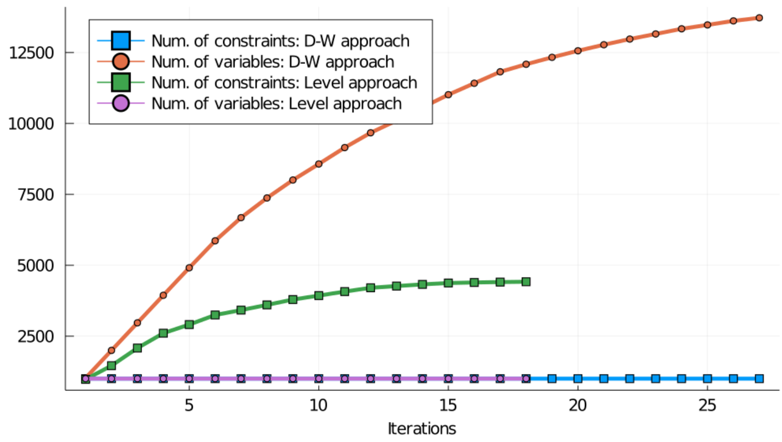


FIGURE 4.5: Size of the master programs of both the Level Method and the Dantzig-Wolfe algorithm on the `FERC 2015-07-01 high wind` instance. The Level Method adds cuts, which implies that the number of constraints is growing. On the other hand, D-W adds columns, which implies that the number of variables is growing. The robustness of the Level Method translates into a master program that grows less rapidly than D-W.

which shows how the convergence of LOC over iterates translates to the distance of prices from CHPs). Instead, the Level Method reaches near-optimal prices in fewer iterations. This is an inherent advantage of the Level Method, which is by design a *stabilization* approach.

Finally, we comment on the *primal method* benchmark. The FERC model exceeds what a primal CHP approach such as Hua and Baldick (2017) can model, since it includes ramp constraints and time-dependent startup costs. The integer relaxation is therefore expected to lead to an *approximation* of CHP. The quality of the primal method largely depends on the tightness of the formulation. In this respect, the FERC model is derived from a careful review of the literature dedicated to tight formulations of the unit commitment model (Morales-España et al., 2013; Sridhar et al., 2013). The quality of the model is discussed in details in Knueven et al. (2020), where it is accompanied by computational experiments of its tightness. As observed in Table 4.1, the primal method turns out to provide a close approximation of CHP *on these FERC instances*. Nonetheless, this is not always guaranteed, as we observe in the next test case (Table 4.4), where the primal method leads to an average lost opportunity costs which is $\sim 60,000$€ higher than CHP, for a market of comparable dispatch cost[19].

The test cases analysed so far suggest a promising performance for the Level Method. Nevertheless, even if these FERC instances are of realistic scale insofar as the number of power plants are concerned, we are interested in computing *prices*. This suggests that it is the *dimension of the price space* that matters the most. There are essentially two ways[20] to increase the price dimension: (i) augmenting the *time horizon*—the horizon of future EU markets will be 96 periods of 15 minutes—and (ii) adding a network—which is unavoidable in both the EU and the US markets. This motivates the next test cases.

4.4.2 *EU test cases*

We now extend our analysis to use cases *with a network*. The EU dataset that we utilize is the one used in Aravena and Papavasiliou (2016). The network data is based on Hutcheon and Bialek (2013), and is constructed among others from an ENTSO-E database[21]. The market suppliers are modelled as a slightly simpler version of the UC model than the FERC test case, essentially simplifying the cost structure: there is a single start-up cost, instead of the variable start-up costs of FERC, and the marginal production cost is constant. All the cases are simulated over 6 different load profiles. As we are interested in studying the scalability of the Level Method and D-W algorithm with respect to the *network* and the *time horizon*, the data has been aggregated into two test cases: BE and BE-NL, which are described in Table 4.2. As detailed in section 4.1, Euphemia, the EU market clearing algorithm, currently computes prices for ~ 40 bidding zones, and is expected to move to 15-minute granularity (96 time periods) in the near future. This makes our two tests cases with 96 periods very relevant proxies of the evolving EU context with respect to price dimensionality.

The final results are obtained with the stopping criterion set to 0.01%, as for the FERC cases. Table 4.3 shows the sensitivity of the Level Method towards parameter α, previously discussed in section 4.3.3. Table 4.4 presents results for the BE test case with multiple *time horizons*. Figs. 4.6 and 4.7 illustrate the convergence of the BE test case with 96 periods. Table 4.5 presents a comparison for different *network sizes*. It is worth noting that,

19 We refer the reader to chapter 3 for a more extensive comparison of CHP with ELMP.

20 A third way would be the introduction of reserve. The current EU DA market does not co-optimize energy and reserve, which is why it is not considered in our analysis. Nevertheless, art. 40 of EGBL guidelines indicates that this could constitute a future evolution of the EU market.

21 This is the same as the so-called " CWE dataset" used in chapter 3.

Test case	Bidding Zones	Lines	Generators
BE	30	30	74
BE-NL	59	63	145

TABLE 4.2: Description of the size of the EU instances.

α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Level iter	54	44	45	43	41	43	45	48	60

TABLE 4.3: Sensitivity of the Level Method with respect to parameter α on the BE 96-period case (average over 6 instances).

in all the test cases (except the 12-period BE test case, which is however less relevant for practical applications), the Level Method turns out to be superior to D-W in terms of iteration count. Furthermore, we observe that the benefits of the Level Method are magnified when increasing the dimension of the price space.

More specifically, insofar as sensitivity with respect to the *time horizon* is concerned, Table 4.4 demonstrates that the Level Method scales well with respect to the horizon of the problem as it increases from 19 to 44 iterates as the horizon grows from 12 to 96 periods. On the other hand, the performance of D-W is seriously harmed by the increase of the horizon: the number of iterates increases from 19 to 236. The stable behavior of the Level Method is corroborated by Fig. 4.6. We observe that, within 6 iterates, it already reaches a price that achieves lower LOC than IP pricing. Fig. 4.7 also presents the convergence of both algorithms on the same instance in terms of the price distance to the optimum. Being capable to reach quickly decent price candidates is an attractive feature for the EU implementation of CHP, recalling from section 4.1 that Euphemia is currently granted 12 minutes for computing the EU day-ahead market matchings and prices.

As far as the *network size* is concerned, Table 4.5 presents the sensitivity with respect to the two use cases. Perhaps surprisingly, neither of the methods is strongly affected by the size of the network, rather the contrary. On the instance with 24 periods, the benefits of the Level Method are similar as in the FERC cases. On the 96-period instances, the Level Method moves from five to four times faster than D-W on the BE and BE-NL cases, in terms of iteration count. Overall, D-W seems much more affected by the increase in the *time horizon* rather than the presence of a *network*, to which

horizon	12	24	48	96
Dispatch Cost [€]	2,759,706	4,956,513	11,328,351	24,097,373
IP LOC [€]	377,528	146,167	281,649	2,617,852
Primal LOC [€]	50,871	64,323	83,172	98,391
CHP LOC [€]	7,237	11,905	21,745	31,403
Level iter	19	26	32	44
Level av. time/iter[a][s]	0.5 (0.05)	0.8 (0.1)	2.0 (0.4)	5.8 (1.6)
Level total run time [s]	10	21	65	255
D-W iter	19	40	77	236
D-W av. time/iter[a][s]	0.4 (0.02)	0.7 (0.1)	1.9 (0.3)	6.9 (2.1)
D-W total run time [s]	7	27	146	1622

[a] $(\cdot)$ denotes the average time per iterate for solving the "master programs" (i.e. master plus projection in the case of the Level Method).

Table 4.4: Results of the Level Method and the Dantzig-Wolfe algorithm on the BE test case (average over 6 instances).

horizon	24		96	
test case	BE	BE-NL	BE	BE-NL
Level iter	26	21	44	42
Level av. time/iter[a][s]	0.8 (0.1)	1.5 (0.3)	5.8 (1.6)	12.3 (4.8)
Level total run time [s]	21	31	255	514
D-W iter	40	32	236	156
D-W av. time/iter[a][s]	0.7 (0.1)	1.3 (0.2)	6.9 (2.1)	12.7 (3.7)
D-W total run time [s]	27	42	1622	1976

[a] $(\cdot)$ denotes the average time per iterate for solving the "master programs" (i.e. master plus projection in the case of the Level Method).

Table 4.5: Results of Level Method and the Dantzig-Wolfe algorithm for different network sizes (average over 6 instances).

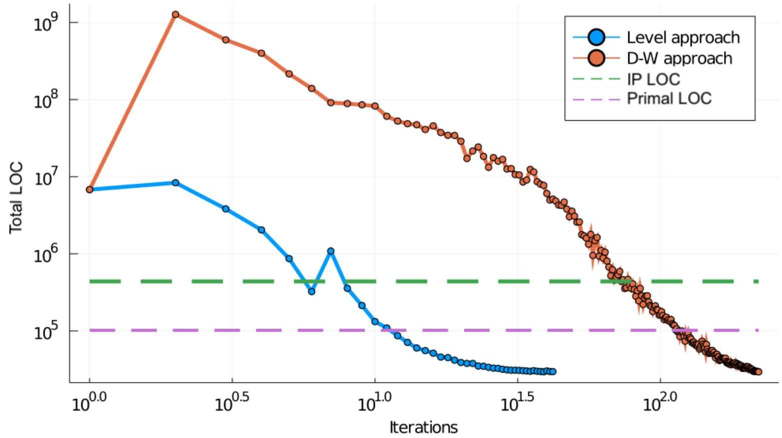


FIGURE 4.6: Convergence of the Level Method and the Dantzig-Wolfe algorithm, measured by the lost opportunity costs (IP pricing and the primal method are used as benchmark thresholds), on the BE summer weekday 96-periods instance. Both axes are in logarithmic scale.

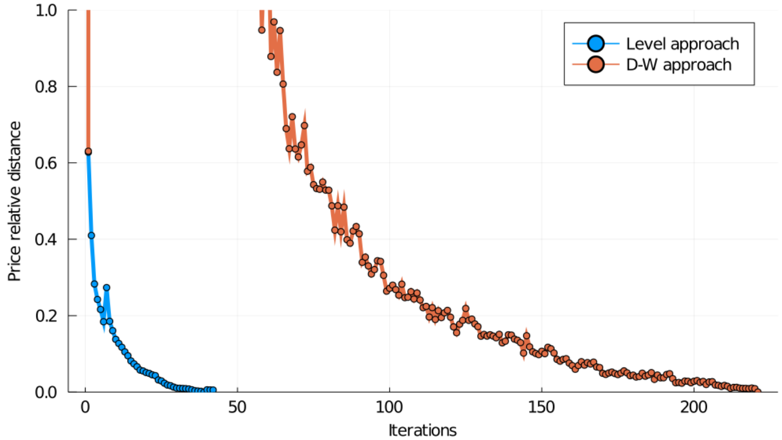


FIGURE 4.7: Convergence of the Level Method and the Dantzig-Wolfe algorithm, measured by the price relative distance to CHP, on the BE summer weekday 96-periods instance.

the test cases suggest D-W is rather robust. This is possibly due to the fact that D-W is required to explore in the space of promising power plant schedules—and these schedules become more and more numerous when increasing the horizon—while the network size does not affect immediately the number of schedules.

It should be stressed that the aforementioned computational gains can make a difference for the practical implementation of CHP, keeping in mind the 12-minute run time limit of Euphemia[22]. From Table 4.4, we observe that the Level Method requires less than 5 minutes on average for solving a 96-period instance. The D-W algorithm requires 27 minutes.

The computational times reported in our results may of course not be representative of the implementation of the EU NEMOs, as solving the slaves in parallel and increasing the computational power would reduce the run time. Assuming an idealized parallelization of the slaves—which is very optimistic considering the NEMOs currently run Euphemia on 8 threads (NEMO Committee, 2019)—, the run time per iterate would be lower-bounded by the time for solving the master programs (master plus projection programs for the Level Method, as reported between brackets in the tables). As an example, the "most difficult" BE-instance was solved in 266 iterates by D-W, with 2.3 sec/iter for solving the master program. This implies a lower bound of more than 10 minutes for obtaining the CHP. On the same instance, the Level Method required 37 iterates, with 1.4 sec/iter for solving the masters, which amounts to a total of less than 1 minute. Furthermore, whereas the price dimension of our test cases has been selected so as to be comparable to the EU market, the number of generators (or market bids) is well below the value that occurs in practice. As an order of magnitude, Euphemia currently solves instances with around 160,000 hourly orders (convex) and 4,000 block orders (non-convex) (NEMO Committee, 2020a). This suggests that the time for solving the master programs would likely be higher on the real instances of Euphemia.

4.5 CONCLUSION

This chapter proposes a bundle stabilization approach for efficiently solving convex hull pricing. We demonstrate that the Level Method is able to converge within a few iterations to a certain target gap, while exhibiting a stable behaviour, on large instances which, in terms of price space dimension, are comparable to the size of the EU day-ahead auction.

It is likely that the choice of the best algorithm for solving CHP will depend on the specific use-case: the dimension of the network, the time

22 See footnote 7.

horizon, the complexity of the unit commitment or market orders, the run time that is afforded to the algorithm, etc. Although no method can conceivably provide an ultimate solution for computing CHP in an arbitrarily complex setting, the Level Method indicates the promising behaviour of a family of "bundle approaches". This suggests areas of future research on alternative bundle approaches, such as the *Proximal Stabilization* method, the *Doubly-Stabilized Bundle Method* (Frangioni, 2020) or the *Boxstep method* (Marsten et al., 1975), which appear to be well suited for solving the CHP Lagrangian relaxation.

Another question for future research relates to how the proposed approach can be adapted in case one of the following assumptions is relaxed: the convexity of the grid model and the separability of the suppliers' profit maximization problems.

Having scalable algorithms capable to compute CHP on large instances also enables more extensive quantitative analysis of its economical behaviour, as exemplified by the developments of chapter 3. As far as the EU market is concerned, future works could (i) expand the tests on realistic instances of Euphemia—our preliminary tests show that the Level Method can solve the 4MMC[23] run of Euphemia in ~ 1 minute—, (ii) examine the effects of non-uniform pricing on enhancing welfare in the EU day-ahead market, and (iii) understand distributional effects of non-uniform pricing as well as gaming effects.

23 The so-called "4MMC region" refers to a relatively small region (Czech Republic, Slovakia, Hungary and Romania) that Euphemia used to solve in a separate run from the rest of Europe, although it was later merged in 2022 with the rest of Europe in a single run.

Technology	N	MC_g [€/MWh]	SC_g [€]	Q_g^{min} [MW]	Q_g^{max} [MW]
Smokestack	3	3	53	0	16
High Tech	2	2	30	0	7
Med Tech	1	7	0	2	6

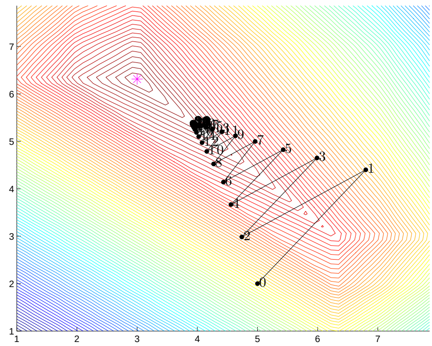
Table 4.6: Data for the stylized example with 6 suppliers (3 "Smokestack" plants, 2 "High Tech" plants and 1 "Med Tech" plant). The data include the number of units, the marginal production cost, the start-up cost, the minumum and maximum output of the plant. It is a modified version of an example proposed by Scarf (1994).

4.A Appendix: Illustration on a 2-D example

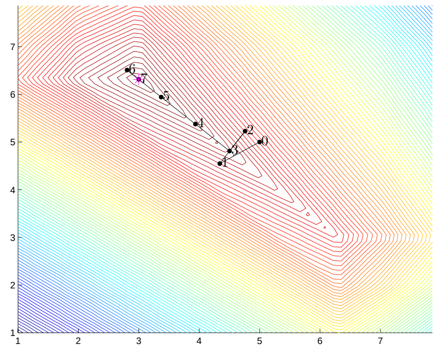
Section 4.3 discusses, among others, four main algorithms to solve the Lagrangian relaxation implied by convex hull pricing: the subgradient method, the EPSD, the Kelley algorithm and the Level Method. These algorithms have been implemented in an earlier version of the work of this chapter (Stevens, 2016). Some of the trade-offs, observations and conclusions that we have reached when implementing these various approaches, and that led us towards the Level Method, are summarised in section 4.3. The objective of this appendix is to provide the reader with a numerical example in two dimensions, which can fruitfully be read in parallel with the discussion of section 4.3 in order to illustrate the "trade-offs, observations and conclusions" that are more extensively covered in Stevens (2016).

The aforementioned four algorithms have been implemented on an example with a single node, two hourly periods and an inelastic demand D of, respectively, 30 and 40 MW in periods 1 and 2. There are six suppliers described in Table 4.6. Since there are two periods, the Lagrangian function has two dimensions, that correspond to the prices of periods 1 and 2. Figure 4.8 compares the sequence of iterates generated by each of the four algorithms on this example. The x- and y-axis correspond to the prices in period 1 and 2 (π_1 and π_2). The figures presents the map of the Lagrangian function. The initial price iterate is marked as 0 on each figure. The subsequent numbers correspond to the next iterates computed by each algorithm. The domain is bounded such that $1.5 \leq \pi_t \leq 7$ (i.e. the box Q described in section 4.3). As observed on the figure, the Lagrangian function is piecewise linear with an optimum marked by a star, corresponding to the convex hull prices $\pi = (3,\ 6.3)$.

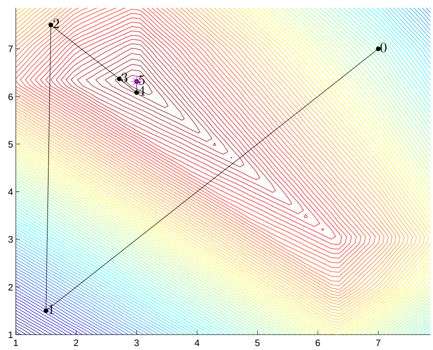
As explained in section 4.3, we observe in Figure 4.8a the typical oscillation behaviour of a subgradient algorithm on a piecewise linear function.



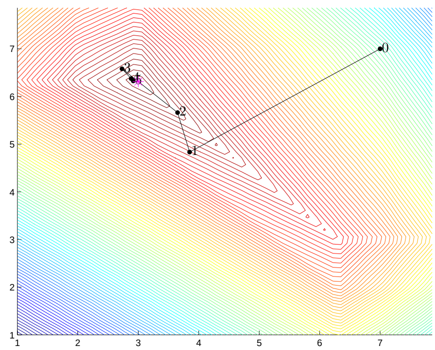
(a) Subgradient path.



(b) EPSD path.



(c) Kelley path.



(d) Level path.

FIGURE 4.8: Iterates sequence of the four algorithms on the modified Scarf example. x- and y-axis correspond to the prices π_1 and π_2.

The algorithm tends to oscillate between two facets of the Lagrangian function, around an edge.

The idea of the EPSD algorithm, summarized in section 4.3, is to "catch the edge" and to follow it, as illustrated on Figure 4.8b. As noticed earlier, the EPSD algorithm can lead to numerical issues since "following an edge" implies to evaluate whether an iterate is still on the edge or not (for instance, determining whether point 5 of Figure 4.8b is still on the edge) which can be numerically challenging. More importantly, we also observe that, as discussed in Wang et al. (2013a,b), computing the direction of the new iteration which is implied by the existing edge requires to retrieve *all* the optimum integer solutions of each supplier profit maximization problem. This can be done using the `populate` routine in `cplex`. However, this requires to explore the whole the branch and bound tree of each profit maximization, in order to retrieve not *one* but *all* the optimum dispatches given a price, thus implying a significant computational burden.

As observed in Figure 4.8c, Kelley's algorithm permits a significant improvement. However, we noted in section 4.3 that the algorithm tends to become unstable in higher dimension, arguing that "this is due to the unstable nature of piecewise linear functions: adding a new supporting hyperplane can move the optimum far from the previous point (i.e. to a corner of the box Q)." This is apparent in Figure 4.8c: the sequence of iterates performs large swings between the corners of the box Q. If this has a mild effect in low dimension, it gives a glimpse to the reader of what may happen if the dimension of the box Q increases to more than 5000, instead of 2, as in the BE-NL 96 periods case of section 4.4.2.

Finally, Figure 4.8d shows how the Level Method tends to stabilize the iterate sequence generated by Kelley, significantly mitigating the oscillations. This asset is what makes the Level Method superior in higher dimension.

Part II

Non-Convexities in the *Long-Term*

5

INDIVISIBILITIES IN INVESTMENT AND THE ROLE OF A CAPACITY MARKET

ABSTRACT. *The topic of pricing non-convexities in power markets has been explored vividly in the literature and among practitioners for the past twenty years. The debate has been focused on indivisibilities in short-term auctions, the computational tractability of some pricing proposals, and the economic analysis of their behavior. In this chapter, we analyse a source of non-convexities that is not discussed as broadly: the indivisibilities in investment decisions. The absence of equilibrium that we are primarily concerned about is the long-term equilibrium. We derive a capacity expansion model with indivisibilities and we highlight the issues arising from it. We discuss its relevance and address one particular argument for neglecting indivisibilities in investment, namely market size. We investigate to what extent a capacity market that clears discrete offers can mitigate the lumpiness problem. We particularly introduce the novel concept of convex hull pricing for capacity auctions. We illustrate the main findings with a numerical experiment conducted on the capacity expansion model used by ENTSO-E to assess the adequacy of the entire European system.*.*

KEYWORDS. Pricing indivisibilities · Investment problem · Capacity market · Convex hull pricing

JEL CLASSIFICATION. C61 · D41 · D44 · D47 · D50 · L51 · Q41

* The chapter reproduces, with minor changes, the content of Stevens et al. (2024b).

5.1 INTRODUCTION

THE restructuring of the electricity industry is work in progress for more than 25 years. Early discussions mainly concentrated on decentralized versus centralized organizations of the market (Stoft, 2002). The latter system emerged, and with it the idea of a centralized market clearing, and a dispatch associated to the so-called merit order of plants that reflects fuel costs. It was quickly recognized that generation plants are also characterized by "indivisibilities", such as start-up cost or minimum time between shutdown and startup. Accounting for these aspects required replacing the merit order-based dispatch by a unit commitment. This invalidated the clean neoclassical interpretation of electricity prices that, according to the doxa, supports competition in the industry. This made the market design more complex and generated a lot of implementation and methodological work, including a vivid debate in the literature and among practitioners about the right way to price in power auctions (O'Neill et al., 2005; Hogan and Ring, 2003).

Indivisibilities also have a long-term dimension. In the same way that plants go through short-run cycles where they are started, operated for some hours, and shut down, they also go through a long-term cycle where they are built, operated over several years, and are eventually dismantled. Each of these stages implies costs that, once incurred, become stranded, hence constituting indivisibilities. In contrast with the short-run market, long-run indivisibilities did not receive much attention, whether in the literature or in practice, so far. A notable exception is Scarf's ground-breaking paper, which recognizes the indivisibilities in the choice of technologies as an issue (Scarf, 1994).

The apparent neglect of indivisibilities in long-term electricity markets contrasts with the attention given to market failures and how these interact with investment incentives. The notion of "missing money" has been central in these discussions since Joskow (2007) enlightened the debate on the subject. The author relies on a stylized capacity expansion model to show that long-term elements are missing in the short-run market, which makes it unable to send adequate investment signals. The missing money debate generated considerable but often inconclusive discussions on the respective merits of different market designs such as energy-only markets and capacity markets. The energy transition in Europe and its implication of fully restructuring the capital stock of the generation system gave a new impetus to the subject. It was recognized that the insufficient incentive to invest was rooted not only in "missing money" but also in missing and incomplete financial markets, which are more difficult issues to explore and remedy. Considerable work has been undertaken in the UK since at least 2013 (UK Department of Energy & Climate Change, 2013; Grubb and Newbery, 2018; Helm, 2017). This led to the idea of using instruments such as contracts for differences (CfDs), power purchase agreements (PPAs) and capacity auctions (CRM) to mitigate this missing incentive for investment (De Maere d'Aertrycke et al., 2017). The general principle of this approach is that existing markets should be complemented by additional market instruments targeted at the incentive to invest. The war in Ukraine and the new European policy of moving away from Russian gas supplies reinforced the sense of urgency of the investment problem. This led to an explosion of papers to remedy not only the impact of high gas prices on the power market but also the possibly insufficient incentives to invest. It reinforced the push for the already mentioned market instruments (CfDs, PPAs, CRM).

This chapter aims at contributing to this literature on investment incentives, but focusing on the—much less discussed—problem of non-convexities in investment. Some papers have focused on the effect of the *short-run* non-convexities on the investment incentives (Mays et al., 2021; Byers and Hug, 2023). Instead, our work focuses on the non-convexities in the investment itself. The goal of an investor to maximize profits still remains the same in the presence of indivisibilities, but these indivisibilities can distort the capacity mix that results from existing incentives. As in Joskow's initial analysis of missing money (Joskow, 2007), we examine the problem through a deterministic capacity expansion model. This corresponds to a complete market (no uncertainty or missing market) and thus makes it possible to focus on the sole effect of indivisibilities. We analyse the possibility that, very much like indivisibilities in the short-run market required a regulatory authority to clear the short-term market, long-term indivisibilities may also require such an authority to coordinate investment. Our formal analysis leads to ideas related to the work of French economists Finon and Roques who claim that there exist fundamental difficulties for coordinating investment in the restructured power market, with the conclusion that direct regulated public intervention should be introduced for that purpose (Roques and Finon, 2017; Finon and Beeker, 2022). The authors refer to this mix of market and public coordination as the "hybrid market"—a notion also supported by Joskow (2022).

In sections 5.2 and 5.3, we introduce a long-term investment model, and we analyse the effects of the market imperfection at work, namely indivisibilities or non-convexities in investment decisions. We show that indivisibilities in investment result in a distortion of incentives for the individual market agents—a long-term *lost opportunity cost*, similarly to what happens in a short-term market with indivisibilities. Is this *long-term* lost opportunity cost important? In section 5.4, we derive one theoretical argument, inspired from the theory of general equilibrium, for neglecting indivisibilities in general, and then discuss its relevance to the investment problem. We show that the issue stemming from discrete investment, under certain pricing approaches, may be arbitrarily large. In principle, the discussion of the first part of the chapter (sections 5.2 to 5.4) applies to any industry. In practice, however, one may expect the problem to be more severe in the electricity sector. Because there are important technical barriers to the storage and transportation of electricity, a shortfall of generation capacity in a given location may not be compensated either by a stock of energy or by raising imports. Unlike many industries, electricity has hardly any means to react to a local shortage of production capacity. These supply and transportation rigidities, combined with an electricity demand which has to be met just in time by production, and which is notably inelastic, especially in the short run, may exacerbate the impact of investment indivisibilities on energy prices—therefore on investment incentives.

The second part of the chapter focuses on the interplay between capacity markets and investment indivisibilities. In section 5.5, we analyse to what extent the long-term lost opportunity cost can be corrected by market mechanisms. Capacity markets are one way to coordinate long-term investments. Some existing capacity markets acknowledge, in their design, the indivisible nature of investment decisions. For example, the Belgian capacity market *only* includes indivisible bids (Elia (2022) art. 235, sec. 6.2). But if the benefits of a capacity market as an instrument to hedge investment risk (De Maere d'Aertrycke et al., 2017) or to mitigate market power in the energy market (Fabra, 2018) have been well analysed (see also Stoft (2002); Cramton and Stoft (2005); Cramton et al. (2013)), little has been said about the effect of the capacity market on the incentives of the agents to invest in a market with long-term indivisibilities. Indeed, if lumpiness of investment has sporadically been mentioned to justify CRMs (Mastropietro et al., 2017), no formal discussion of the argument has been provided so far to the best of our knowledge. We analyse to what extent a capacity market could turn out to be a tool that mitigates the *long-term* lost opportunity costs stemming from indivisibilities, or if it alternatively exacerbates them. We particularly discuss the design of a CRM under discrete offers and we introduce the concept of *convex hull pricing* for capacity auctions. Finally, section 5.6 illustrates our findings with a model of the European system. We perform our simulations with the European Resource Adequacy Assessment (ERAA) model used by ENTSO-E (2021) to estimate the need for investments in Europe.

5.2 THE CONTINUOUS INVESTMENT PROBLEM

The *continuous* investment problem provides us with a useful benchmark[1]. Its analysis was pioneered by Boiteux (1960), who showed that *marginal pricing* provides market agents with the right incentives to invest in the welfare-maximizing generation mix. The analysis resolves the fallacy according to which a peaking unit could not possibly cover its fixed cost by solely relying on market payments. The analysis of Boiteux can be illustrated by considering the following long-term *continuous* investment model (which admits a decentralized interpretation):

$$\max_{q,x,d\geq 0} \sum_{t\in\mathcal{T}} \Delta T_t V_t d_t - \sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}} \Delta T_t MC_g q_{g,t} - \sum_{g\in\mathcal{G}} IC_g x_g \tag{5.1a}$$

$$(\Delta T_t \pi_t) \quad d_t \leq \sum_{g\in\mathcal{G}} q_{g,t} \quad \forall t\in\mathcal{T} \tag{5.1b}$$

$$(\Delta T_t \mu_{g,t}) \quad q_{g,t} \leq x_g \quad \forall g\in\mathcal{G},\ t\in\mathcal{T} \tag{5.1c}$$

$$(\Delta T_t \eta_t) \quad d_t \leq D_t \quad \forall t\in\mathcal{T} \tag{5.1d}$$

The variables x_g, $q_{g,t}$ and d_t stand respectively for the investment in technology $g\in\mathcal{G}$, the actual production from technology g at period $t\in\mathcal{T}$, and the consumption of energy at period t. Investment cost is indicated as IC_g, while marginal cost is indicated as MC_g. The total served demand at period t, d_t, is valued at V_t, which is assumed to be the *right* value of lost load[2] (VOLL). D_t is the observed load while ΔT_t stands for the duration of period t. As indicated by the inequality in the market clearing constraint (5.1b), we assume *free disposal*. The optimality conditions of problem (5.1) are:

$$0\leq q_{g,t} \perp MC_g - \pi_t + \mu_{g,t} \geq 0 \qquad \forall g\in\mathcal{G},\ t\in\mathcal{T} \tag{5.2a}$$

$$0\leq x_g \perp IC_g - \sum_{t\in\mathcal{T}} \Delta T_t \mu_{g,t} \geq 0 \qquad \forall g\in\mathcal{G} \tag{5.2b}$$

$$0\leq d_t \perp -V_t + \pi_t + \eta_t \geq 0 \qquad \forall t\in\mathcal{T} \tag{5.2c}$$

$$0\leq x_g - q_{g,t} \perp \mu_{g,t} \geq 0 \qquad \forall g\in\mathcal{G},\ t\in\mathcal{T} \tag{5.2d}$$

$$0\leq D_t - d_t \perp \eta_t \geq 0 \qquad \forall t\in\mathcal{T} \tag{5.2e}$$

$$0\leq \sum_{g\in\mathcal{G}} q_{g,t} - d_t \perp \pi_t \geq 0 \qquad \forall t\in\mathcal{T} \tag{5.2f}$$

These equations convey three important facts. (i) If a technology is used ($x_g > 0$), then the infra-marginal rents ($\sum_{t\in\mathcal{T}} \Delta T_t \mu_{g,t}$) earned from the short-term market prices π_t by each technology exactly cover the investment cost

1 We repeat here for convenience the classic long-term competitive analysis explained in section 2.6.4.

2 cf. discussion in section 2.4.

IC_g. (ii) This means that long-term profits are zero. (iii) Furthermore, as highlighted by Boiteux, in order for the peaking units (the technology g with the highest MC_g) to recover their fixed costs, there should be at least some hours during which the system is *scarce*, meaning that the demand sets the price ($d_t < D_t$, such that $\pi_t = V_t > MC_{peak}$).

5.3 THE DISCRETE INVESTMENT PROBLEM AND THE *long-term* LOST OPPORTUNITY COST

We now turn to the *discrete* version of model (5.1) that accounts for the lumpiness of investment. Indivisibilities in investment decisions (commissioning or decommissioning) arise naturally from the fact that power plants are large indivisible assets (Williamson, 1966), e.g. nuclear or CCGT plants as well as an offshore wind park are straightforward examples. Indivisibilities also arise indirectly from economies of scale as well as learning effects (Arrow, 1962). "Learning by doing" can be represented as a particular model with indivisibilities (Heuberger et al., 2017) that appears to be of particular interest in certain policy design discussions (Newbery, 2021). The discrete investment model is as follows:

$$z_P^* = \max_{q,x,d} \sum_{t\in\mathcal{T}} \Delta T_t V_t d_t - \sum_{g\in\mathcal{G}} \left(\sum_{t\in\mathcal{T}} \Delta T_t MC_g q_{g,t} + \sum_{i\in\mathcal{I}_g} x_{g,i} IC_{g,i} \right) \tag{5.3a}$$

$$d_t \leq \sum_{g\in\mathcal{G}} q_{g,t} \quad \forall t \in \mathcal{T} \tag{5.3b}$$

$$0 \leq q_{g,t} \leq \sum_{i\in\mathcal{I}_g} P_{g,i}^{max} x_{g,i} \quad \forall g \in \mathcal{G},\ t \in \mathcal{T} \tag{5.3c}$$

$$x_{g,i} \in \{0,1\} \quad \forall g \in \mathcal{G},\ i \in \mathcal{I}_g \tag{5.3d}$$

$$0 \leq d_t \leq D_t \quad \forall t \in \mathcal{T} \tag{5.3e}$$

The investment decisions are modelled with the binary variables $x_{g,i}$. These stand for investment into *lumps* of capacity $P_{g,i}^{max}$ at investment cost $IC_{g,i}$, so that each market agent (or technology) g comes with the set of investment projects $i \in \mathcal{I}_g$. The real-time operations are assumed to be convex[3]. This formulation of the discrete investment problem is similar to the one considered by Scarf (1994) or O'Neill et al. (2005). To ease notation, we

3 The model neglects the non-convexities of short-term production costs, such as start-up or no-load costs, i.e. the short-term variable production cost is thus MC_g. That is, we do not account for the (difficult) problem, tackled in chapters 3 and 4, of how the *short-term* non-convexities should be accounted for in the price signal. Instead, the model of this chapter focuses on the *long-term* non-convexities.

shall denote the total cost of each agent for performing the production plan $(q, x)_g$ as the linear function $c_g((q, x)_g)$ in the remainder of this chapter. Thus, $c_g((q, x)_g) = \sum_{t \in \mathcal{T}} \Delta T_t MC_g q_{g,t} + \sum_{i \in \mathcal{I}_g} x_{g,i} IC_{g,i}$. The (non-convex) production sets defined by constraints (5.3c)-(5.3d) are denoted as $\mathcal{X}_g$, while the (convex) consumption set defined by constraint (5.3e) is denoted as $\mathcal{X}_d$.

The short-term *marginal prices*—or *merit order* prices—are the ones stemming from the market when the investment decisions are fixed. In this chapter, we are particularly interested in finding prices that support the welfare-maximizing investment[4]. We therefore assume throughout this chapter that the installed mix is the optimal investment $x^{**}_{g,i}$, as if a central planner were solving problem (5.3).

Definition 5.1 (Marginal Pricing)**.** *Let x^{**} be the values of the binary variables optimizing problem* (5.3)*. The marginal prices are defined as the dual variables π^M obtained from solving the following (convex) problem, in which the variables x of problem* (5.3) *are fixed to x^{**}:*

$$\max_{d,q} \sum_{t \in \mathcal{T}} \Delta T_t V_t d_t - \sum_{g \in \mathcal{G}} c_g((q, x^{**})_g) \tag{5.4a}$$

$$(\Delta T_t \pi^M_t) \quad \sum_{g \in \mathcal{G}} q_{g,t} \geq d_t \qquad \forall t \in \mathcal{T} \tag{5.4b}$$

$$(q, x^{**})_g \in \mathcal{X}_g \qquad \forall g \in \mathcal{G} \tag{5.4c}$$

$$d \in \mathcal{X}_d \tag{5.4d}$$

The marginal pricing approach captures the two-stage nature of an investment cycle. The supplier first decides on the discrete decision (e.g. investing in a new power unit). The associated fixed cost is then considered as *sunk*. Thus, the price reflects the cost of operating the plant given the fixed discrete decisions (i.e. the short-term so-called *merit order*).

The concern with these marginal prices, compared to marginal pricing in the *continuous* investment problem, is that in general they do not support the optimal investment. This price alone does not support an equilibrium: at the socially optimal investment plan, some agents will have incentives to enter or to leave the market. Intuitively, the lumpiness of investment can make it socially optimal to over-dimension the investments, which in turn keeps the prices too low to render the investment profitable in the first place[5]. This fundamental problem of the discrete investment problem was

4 In other words, the dynamics that take us from a sub-optimal mix to this optimal investment are out of scope for this chapter.

5 A recent example is Finland. In April 2023 Olkiluoto-3, which is a 1,600MW nuclear unit, entered into operation and led to a significant price drop.

highlighted by Scarf (1994) (see also the analysis by Williamson (1966)). Before illustrating it in Example 5.1, we proceed with some definitions that characterise the incentives of the market agents. All suppliers and consumers are assumed to be price-takers and to act so as to maximize their selfish profit.

Definition 5.2 (Supplier Profit)**.** *Agent g is assumed to maximize its selfish profit function $\mathcal{P}_g$, under market price π, which is defined as follows:*

$$\max_{(q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(q,x,\pi) \equiv \sum_{t\in\mathcal{T}} \pi_t \Delta T_t q_{g,t} - c_g((q,x)_g) \tag{5.5}$$

Definition 5.3 (Demand Surplus)**.** *The load is assumed to maximize its selfish surplus function $\mathcal{U}$, under market price π, defined as follows:*

$$\max_{d\in\mathcal{X}_d} \mathcal{U}(d,\pi) \equiv \sum_{t\in\mathcal{T}} \Delta T_t (V_t - \pi_t) d_t \tag{5.6}$$

Definition 5.4 (Competitive Walrasian Equilibrium)**.** *The allocation (q^*, x^*, d^*) together with the market price π constitute a competitive Walrasian equilibrium if*

(i) for each supplier g, $(q^*, x^*)_g$ optimizes its profit maximization problem (5.5) *under price π ; d^* optimizes the load surplus maximization problem* (5.6) *under price π ;*

(ii) the market clears ($\sum_{g\in\mathcal{G}} q^*_{g,t} \geq d^*_t \quad \forall t \in \mathcal{T}$).

Since the market is non-convex, a competitive equilibrium is not guaranteed to exist. Under a *centralized* production and consumption plan (q^*, x^*) and d^*, chosen so that condition (ii) of Definition 5.4 is met, there may be no price that satisfies condition (i). Assuming that the private agents maximize their profit (Definition 5.2 and 5.3), the violation of condition (i) is measured by the *long-term* lost opportunity cost.

Definition 5.5 (Long-term Lost Opportunity Cost)**.** *The lost opportunity cost (LOC) is the difference between the selfish maximum profit if self-scheduling and the as-cleared profit (with allocation (q^*, x^*, d^*)) under price π. For each supplier g, it is expressed as:*

$$0 \leq LOC_g(\pi) = \overbrace{\max_{(q,x)_g\in\mathcal{X}_g} \mathcal{P}_g(q,x,\pi)}^{\text{selfish maximum profit}} - \overbrace{\mathcal{P}_g(q^*,x^*,\pi)}^{\text{as-cleared profit}} \tag{5.7}$$

For the demand, it is expressed as:

$$0 \leq LOC_d(\pi) = \max_{d\in\mathcal{X}_d} \mathcal{U}(d,\pi) - \mathcal{U}(d^*,\pi)$$

This concept has been widely used in the context of pricing non-convexities in power auctions[6]. In an investment context, the long-term lost opportunity cost measures the financial incentives that each profit-maximizing agent has to commission or decommission power plants in a way that deviates from the efficient capacity mix (the one solving problem (5.3)). The LOC could fruitfully be viewed as the sum of two quantities. In some cases, an LOC corresponds to a *shortfall of revenue*. For instance, a new investment that would be socially efficient, while it is unprofitable, implies that the investor would bear a shortfall of revenue. Alternatively, an installed plant that, from a social efficiency viewpoint, should stay in the market although it is unprofitable, would also face a shortfall of revenue. In a capital-intensive industry such as power production[7], a revenue shortfall stands for a threat of not recovering investment cost. In other cases, LOC corresponds to a *foregone opportunity*. For instance, an investor who, from a social efficiency viewpoint, should restrain from investing, while his investment project is profitable, would forego an opportunity. Alternatively, if it would be socially efficient to retire an existing plant, although it is profitable, then the owner would also forego an opportunity. Mathematically, the revenue shortfall (RS_g) and the foregone opportunity (FO_g) can be related as follows to the definition of LOC (cf. Figure 5.1). Looking at the two terms of expression (5.7), there are three cases (by definition, $\max_{(q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(q,x,\pi) \geq \mathcal{P}_g(q^*,x^*,\pi)$):

(A) Either $\mathcal{P}_g(q^*,x^*,\pi) \geq 0$, in which case there is no revenue shortfall, and the LOC is a "foregone opportunity" ($LOC_g = FO_g$), i.e. the investor does not loose money, but he could gain more by deviating from the socially efficient plan;

(B) Or $\mathcal{P}_g(q^*,x^*,\pi) < 0$. In this case, there are two alternatives: (B1) Either $\max_{(q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(q,x,\pi) \leq 0$, then the LOC is a revenue shortfall ($LOC_g = RS_g$); (B2) Or $\max_{(q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(q,x,\pi) > 0$. In this case, the LOC can be equivalently written as the following sum: $LOC_g(\pi) = [\max_{(q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(q,x,\pi) - \mathcal{P}_g(0,0,\pi)] + [\mathcal{P}_g(0,0,\pi) - \mathcal{P}_g(q^*,x^*,\pi)] = FO_g + RS_g$.

As highlighted in case (B2), the revenue shortfall may fruitfully be viewed as a specific "lost opportunity", in which the as-cleared profit is negative

6 Cf. chapter 3.

7 For a peaking unit operating a few hours per year, or for an offshore wind park, the investment cost stands for most of the total cost of the asset. For a mid-load gas-fired CCGT plant with an annualized investment cost of $\sim 80,000$€/MW/y and a production cost of ~ 50€/MWh with a capacity factor of ~40%, the investment cost would stand for ~30% of its total cost.

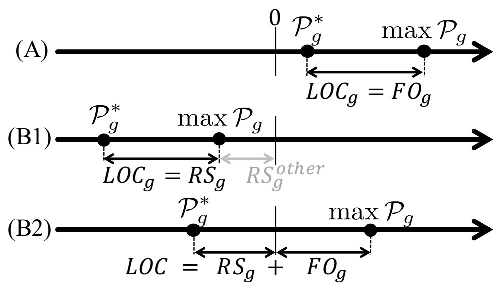


FIGURE 5.1: Graphical illustration of the relationship between LOC, RS and FO. $\mathcal{P}_g^*$ and $\max \mathcal{P}_g$ denote, respectively, the as-cleared profit and the maximum profit.

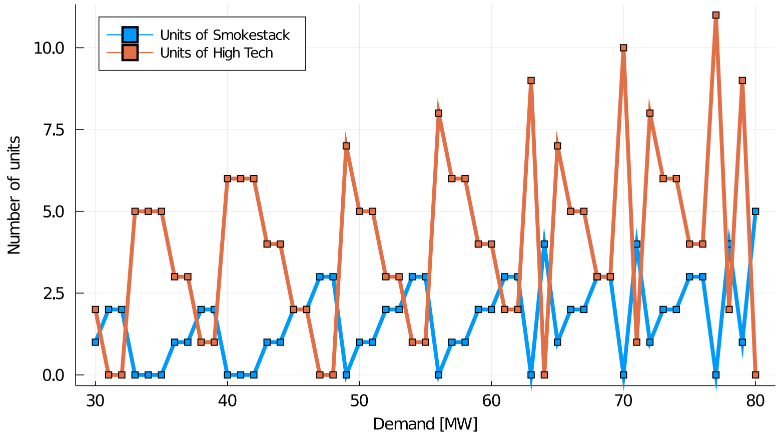
Technologies	Capacity [MW] (P^{max})	Investment Cost [€/unit] (IC)	Marginal Cost [€/MWh] (MC)
Smokestack	16	53	3
High Tech	7	30	2

TABLE 5.1: Power plant data in Scarf's example (Scarf, 1994).

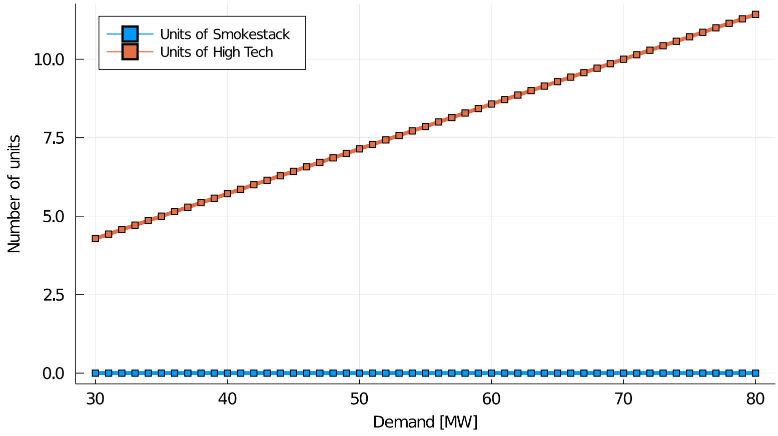
and the "opportunity" is not to invest ($x^* = 0$ and $\mathcal{P}_g(0, 0, \pi) = 0$)[8]. We shall reuse the notions of RS_g and FO_g in the sequel, especially in section 5.6.

Example 5.1. *Consider the classic example proposed by Scarf (1994). This can be described as a discrete investment problem into two different technologies (Table 5.1). One technological option is Smokestack, the other is High Tech plants. A central planner solves problem* (5.3) *in order to determine the cost-minimizing number of power plants of each technology to install so as to meet the perfectly inelastic demand D. Figure 5.2a reports the cost-minimizing investment choices as a function of load. The lumpiness of investment translates into highly fluctuating investment decisions, depending on market demand. For the sake of comparison, Figure 5.2b illustrates what would be the optimal expansion if the investment decisions were continuous. Since the average cost of the Smokestack plant is 6.3125€/MWh, while it is 6.2857€/MWh for the High Tech plant, only High*

8 In case (B1), $\max_{(q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(q, x, \pi) < 0$ may happen if possibility of inaction does not hold, i.e. if there are barriers of exit. In this case, an investor may have a shortfall of revenue, without having the possibility to act differently ($LOC_g = 0$). Accounting for this subtlety requires to introduce a distinction between a "revenue shortfall that expresses a *lost opportunity*" (RS_g) and a "revenue shortfall that is due to a *barrier of exit*" (RS_g^{other}).



(a) Discrete investment



(b) Continuous investment

FIGURE 5.2: Welfare maximizing investment decisions under discrete and continuous investment as a function of the load.

Tech plants would have been built, so as to precisely meet demand (recall that the example assumes a constant uniform demand D). Let us now consider the case in which the demand equals 60MWh. The optimal investment is to build 2 Smokestack plants and 4 High Tech plants. Under this allocation, the marginal price is 3€/MWh, which corresponds to the marginal cost of the Smokestack plants. The two suppliers face a lost opportunity cost (in this case, a revenue shortfall) of 106€ for the Smokestack plants and 92€ for the High Tech plants. By contrast, under continuous investment, the LOCs are zero for both technologies.

5.4 THE THEORETICAL MAGNITUDE OF LOST OPPORTUNITY COSTS

Indivisibilities in investment have sometimes been overlooked on the basis that, when the market size increases, "inefficiency caused by the lumpiness of generators is negligible". As Stoft argues: "this impact of lumpiness is dramatic, but it occurs in an unrealistically small market. [...] This inefficiency declines in proportion to the size of the market." (Stoft, 2002, pp. 130–131). In other words, the effect of indivisibilities may be dramatic in Example 5.1, but it tends to vanish when the size of the system increases. The same reasoning is supported by Byers and Hug (2023). In this section, we assess whether this might be theoretically true. The intuition that non-convexities would smooth out when the market size increases rests on solid theoretical foundations. A strong result was provided in the late sixties, in the theory of general equilibrium, by Arrow and Starr[9], in order to justify the crucial assumption of convexity that is needed for ensuring the *existence* of a competitive equilibrium. We briefly state the result of Arrow and Starr, before showing how it can be adapted to our problem statement. We then use this result in order to first derive a positive result, and then a more negative result. The settings considered by Arrow and Starr differ in two manners from our settings of section 5.3: (i) their *pricing rule* differs from marginal pricing, and (ii) the *metric* that they use for measuring the distance from competitive equilibrium differs from LOC.

As far as the pricing scheme is concerned, in the absence of competitive prices, the question of what will be the price that prevails in the non-convex market remains open. An alternative to marginal pricing (Definition 5.1), consists of computing the prices from the *closest convex economy* in which a competitive equilibrium exists. Mathematically, the closest convex economy means the convex relaxation of problem (5.3) in which the production sets ($\mathcal{X}_g$) are replaced by their *convex hull* ($\text{conv}(\mathcal{X}_g)$).

9 Starr (1969) shows the result for a pure exchange economy, while Arrow and Hahn (1971) show it for a more general case of an economy that includes non-convex production.

Definition 5.6 (Convex Hull Pricing). *The convex hull prices π^{CH} are defined as the dual variables obtained from solving the following convex problem:*

$$z_D^* = \max_{d,q,x} \sum_{t\in\mathcal{T}} \Delta T_t d_t V_t - \sum_{g\in\mathcal{G}} c_g((q,x)_g) \tag{5.8a}$$

$$(\Delta T_t \pi_t^{CH}) \quad \sum_{g\in\mathcal{G}} q_{g,t} \geq d_t \qquad \forall t \in \mathcal{T} \tag{5.8b}$$

$$(q,x)_g \in \operatorname{conv}(\mathcal{X}_g) \qquad \forall g \in \mathcal{G} \tag{5.8c}$$

$$d \in \mathcal{X}_d \tag{5.8d}$$

Although they do not use this nomenclature, this is the pricing approach assumed by Arrow and Starr[10]. Let us notice that the investment costs $IC_{g,i}$ appear in problem (5.8) through the function $c_g((q,x)_g)$, while they are not present in the marginal pricing problem (5.4), since the investment decisions are fixed.

Regarding the metric used for measuring the distance to an equilibrium, there are two options that are worth examining: either condition (ii) of Definition 5.4 holds—or is enforced—and (i) is violated; or condition (i) holds, in which case (ii) is violated. The first case corresponds to what has been considered in section 5.3, in which distance to the equilibrium is measured by the LOC. The setting analysed by Arrow and Starr corresponds to the second case. It can be viewed as a purely *decentralized* setting: the producers and consumers leave or enter the market depending on the price they observe, in a manner that satisfies (i). Then, the discrepancy between demand and production—the violation of condition (ii)—is measured by the *social excess demand*.

Definition 5.7 (Social Excess Demand). *Let $q_{g,t}^\dagger$ and $d_t^\dagger$ be decentralized production and consumption plans of the private agents under price π, respecting condition (i) of Definition 5.4. The social excess demand (SED) is defined as:*

$$SED(q^\dagger, d^\dagger) = d^\dagger - \sum_{g\in\mathcal{G}} q_g^\dagger. \tag{5.9}$$

5.4.1 *Convex hull pricing with decentralized decisions*

We first consider the same setting as that assumed in the work of Arrow and Starr. Let $((q^*, x^*, d^*), \pi^{CH})$ be the equilibrium in the *closest convex economy*,

10 This happens to also be a pricing proposal that has been advocated—for other reasons—by Hogan and Ring (2003) for pricing in non-convex power auctions under the name Convex Hull (CH) Pricing. Taking the convex hull of the constraints (5.3c)-(5.3d) amounts to solving the Lagrangian relaxation of the problem in which constraint (5.3b) is relaxed. The convex hull prices then correspond to the associated Lagrangian multipliers. This connection with the Lagrangian dual problem justifies the label z_D, for denoting the *dual* objective value.

i.e. the solution of problem (5.8). The allocation (q^*, x^*, d^*) can, in general, be infeasible. Therefore, we shall seek an allocation $(q^\dagger, x^\dagger, d^\dagger)$ that solves problems (5.5) and (5.6) under price π^{CH} (condition (i) in Definition 5.4 is met), even if it does not clear the market (condition (ii) in Definition 5.4 can be violated). How would this mismatch between supply and demand grow with the market size?

Example 5.2. *Consider an investment problem with one single power plant technology with the following characteristics: the investment cost is 50€/MWh, the production cost is 10€/MWh and the indivisible size of the power plant is 100MW (i.e. investing in one plant costs 100MW×50€/MWh= 5,000€/h). Let us assume that one can invest in any non-negative integer number of power plants. We also assume that there is a single period and that the VOLL is equal to 1,000€/MWh. If the demand is $D = 250$MWh, the optimal allocation in the convex hull of this economy is $x = 2.5$ and $q = D = 250$, for which the convex hull price is $\pi^{CH} = 60$€/MWh. At this price, each plant is indifferent between either producing (and investing) zero, or producing at 100MWh (both production plans lead to a zero profit). There exists a decentralized decision to construct two power plants so as to produce 200MWh. On the other hand, at this price, the demand is willing to consume 250MWh. Thus, the social excess demand is 50MWh.*

Intuitively, if the market grows (the demand D increases), the social excess demand will always be bounded by 50 MWh (which can be viewed as a measure of the non-convexity of the production set). This intuition is formally stated and proven to hold for a general case in Proposition 5.1, which is the translation of the Theorem of Starr and Arrow to our problem[11].

Proposition 5.1. *Let π^{CH} denote the convex hull prices and (q^*, x^*, d^*) the associated allocation in the convex problem, where both are obtained from solving problem (5.8). Then, there exists an allocation $(q^\dagger, x^\dagger, d^\dagger)$ such that*

(i) $(q^\dagger, x^\dagger)$ *solve problem* (5.5) *under price* π^{CH}

(ii) $d^\dagger$ *solves problem* (5.6) *under price* π^{CH}

(iii) The difference of social excess demand is bounded[12]

$$|SED(q^\dagger, d^\dagger) - SED(q^*, d^*)| = |(d^\dagger - \sum_g q_g^\dagger) - (d^* - \sum_g q_g^*)| \leq \sqrt{|\mathcal{T}|}A$$

11 Note that $r(\cdot)$ denotes the *inner radius* of a set (the definition is recalled in Appendix 5.A). To provide an intuition to the reader, in the previous example, $r(\mathcal{X}_g) = 50$. Indeed, the set of possible investments is $\{0, 100\}$, meaning a ball of radius 50 MW spans any $x \in [0, 100]$.

12 Note that if $\pi_t > 0\ \forall t \in \mathcal{T}$, then we deduce that $d^* = \sum_g q_g^*$ and the expression becomes $|d^\dagger - \sum_g q_g^\dagger| \leq \sqrt{|\mathcal{T}|}A$.

with $A \geq r(\mathcal{X}_g)\ \forall g$.

The proof, largely inspired from the one of Arrow and Hahn (1971) that we adapt to our problem statement, is provided in the appendix 5.A of this chapter (which also contains all the other proofs of the chapter). The Proposition shows that, under these assumptions of price and metric, the discrepancy between supply and demand, caused by the indivisibilities, is *bounded*. The bound depends upon the number of commodities that are exchanged as well as the measure of non-convexity of each production set, *but it is independent of the size of the market*. If the number of consumers and suppliers is multiplied, while keeping similar production sets, the bound remains unchanged, meaning that its ratio relative to the size of the market tends to zero.

5.4.2 *Convex hull pricing with centralized decisions*

Let us now assume that the production and the consumption plans are decided by a central planner so that the market clears and the solution maximizes social welfare. Let (q^{**}, x^{**}, d^{**}) be the welfare-maximizing allocation, obtained from solving problem (5.3). The market price is again assumed to be the convex hull price π^{CH}. Under this setup, condition (ii) in Definition 5.4 is met, while condition (i) is violated (the violation being measured by the LOC).

Example 5.3. *We consider the same data as in Example 5.2. If* $D = 250MWh$*, the welfare maximizing allocation is* $x = 3$ *so that* $q = D = 250$ *MWh. The convex hull price is* $\pi^{CH} = 60$*€/MWh. At this price, the non-constructed power plants face an LOC of 0€. Two of the constructed power plants—the ones producing 100 MWh each—face an LOC of 0€. The plant at the margin, producing 50 MWh, faces a loss of 2,500€.*

Intuitively, if the market grows (D increases), there will always be one single frustrated plant at the margin, which faces a revenue shortfall of at most 5,000€. This intuition is formally stated and proven for a general case in the following Proposition[13].

Proposition 5.2. *Let* (q^{**}, x^{**}, d^{**}) *be the welfare maximizing allocation, obtained from solving problem* (5.3). *Let* π^{CH} *denote the convex hull prices, obtained from solving problem* (5.8). *Then, the total lost opportunity cost is bounded:*

$$\sum_{g \in \mathcal{G}} LOC_g(\pi^{CH}) + LOC_d(\pi^{CH}) \leq \rho|\mathcal{T}| \tag{5.10}$$

13 A similar proposition is also provided by Chao (2019), although the proof proposed in Appendix 5.A is different.

with $\rho = \max_{g\in\mathcal{G}} \rho_g$ *and* ρ_g *defined as follows:*

$$\rho_g = \max_{(\hat{q},\hat{x})_g \in \text{conv}(\mathcal{X}_g)} \left\{ \hat{c}_g(\hat{q},\hat{x}) - c_g(\hat{q},\hat{x}) \right\} \tag{5.11}$$

$$\hat{c}_g(\hat{q},\hat{x}) = \min_{\substack{(q,x)_g \in \mathcal{X}_g \\ q_{g,t} \geq \hat{q}_{g,t}}} c_g((q,x)_g) \tag{5.12}$$

Let us notice that, in Example 5.3, $\rho_g = 5,000$€. Indeed, $|\mathcal{T}| = 1$ and a worst cost increment of 5,000€ could occur if a plant is asked to produce ϵ (the convex hull allocation is $x^*_g = \epsilon/100 \approx 0$, while a feasible allocation requires to build an entire power plant, $x_g = 1$, which comes at a cost of 5,000€). Similarly to Proposition 5.1, the bound *does not depend on the market size*. If the market grows (increasing the load and the number of suppliers with similar production sets $\mathcal{X}_g$) in such a way that $z^*_P \to \infty$, since the total LOC remains bounded, its ratio with respect to the market size tends to zero, i.e. $LOC(\pi^{CH})/z^*_P \to 0$. In other words, under convex hull prices, the lost opportunity costs do not spread over the entire market but remain contained to a small number of plants at the margin.

5.4.3 *Marginal pricing with centralized decisions*

We now turn to the configuration considered in section 5.3. The production and the consumption plans are decided by a central planner but the market prices are the marginal—merit-order—prices π^M, as computed from problem (5.4).

Example 5.4. *We consider once more the same data as in Example 5.2. For* $D = 250$ *MWh, the social welfare maximizing allocation is* $x = 3$*, so that* $q = D = 250$ *MWh.* π^M *=10€/MWh. Under this price, all the plants that are not constructed are in equilibrium. But each of the three constructed plants faces a revenue shortfall of 5,000€—and not only the plant at the margin—for a total of 15,000€.*

Intuitively, on this stylised example, when the demand D increases, the number of new plants constructed at a loss increases, and so does the lost opportunity cost which *does grow* with the size of the economy[14].

14 The reader may also want to consider these results the other way around. Let us assume that, in the same stylized example, due to some technological improvements, the plants are now available in lumps of 30 MW instead of 100 MW. The non-convex suppliers have been cut into smaller pieces so that the production sets are "less non-convex" than they used to be (as measured with the inner radius). How does the technological change affect the LOC? The reader can verify that, for D = 250 MWh, under convex hull pricing, the LOC is almost divided by three: 1,000€ (compared to the 2,500€ before the technological change). Instead, under marginal pricing, the LOC is 13,500€. This point becomes relevant as we consider distributed resources with smaller capacities in future power grids.

Proposition 5.3. *Let N be the number of times that the input of the market defined in problem* (5.3) *is duplicated, i.e. duplicating N times the set of suppliers* $\mathcal{G}$ *and the load. Let* (q^{**}, x^{**}, d^{**}) *be the associated welfare-maximizing allocation and let* π^M *be the associated marginal prices. Then, in general, the lost opportunity cost is not guaranteed to be bounded, i.e. it may be that* $\lim_{N\to\infty} LOC(\pi^M) = \infty$.

Under marginal pricing, the market failure originating from indivisibilities *could* be arbitrarily large. We stress that Proposition 5.3 does not establish that the LOC grows to infinity *in all cases*, but simply that, in general, it is not guaranteed to be bounded, as opposed to Proposition 5.2. This result highlights that Propositions 5.1 and 5.2 are highly dependent on the pricing scheme that is assumed to hold in the non-convex market. Thus, under alternative prices, indivisibilities *do not* smoothen out and may have a significant impact, even in a large market. In the context of discrete investment, convex hull pricing receives a less intuitive explanation than does marginal pricing. If marginal pricing can indeed be viewed as the classic *merit order* pricing that prevails, then the LOC stemming from indivisibilities is not necessarily expected to vanish when considering a larger market size. The impact described in Proposition 5.3 is arguably exacerbated in Example 5.4 by the fact that there is a *single* peaking technology. The magnitude of the LOC under merit-order pricing will however be studied in a larger system in section 5.6. In the meantime, we turn once again to the *two*-technology example of Scarf, Example 5.1, which illustrates Propositions 5.2 and 5.3, before discussing a possible solution to these lost opportunity costs in section 5.5.

Example 5.5. *We have shown in Example 5.1 that the marginal price is 3€/MWh for a load of* $D = 60$ *MWh, leading to a total lost opportunity cost of 198 €. Instead, for the same load, the convex hull price is 6.2857 €/MWh. At this price, the Smokestack and High Tech technologies face an LOC of 0.857€ and 0€ respectively. As far as the results of this section are concerned, the key observation is Figure 5.3, which reports the total LOC under both pricing schemes for various load scenarios. As expected from Propositions 5.2 and 5.3, the lost opportunity cost grows with the market size under marginal pricing, while it remains bounded under convex hull pricing. The bound, computed using Proposition 5.2, is 53€.*

5.5 CAPACITY MARKETS

What is broken by the presence of indivisibilities in the investment decisions is the possibility to achieve a perfect coordination of private agents solely by means of a uniform energy price signal (Scarf, 1994). A decentralized energy-only market does not guarantee a welfare-maximizing investment. This motivates a policy intervention for coordinating investments.

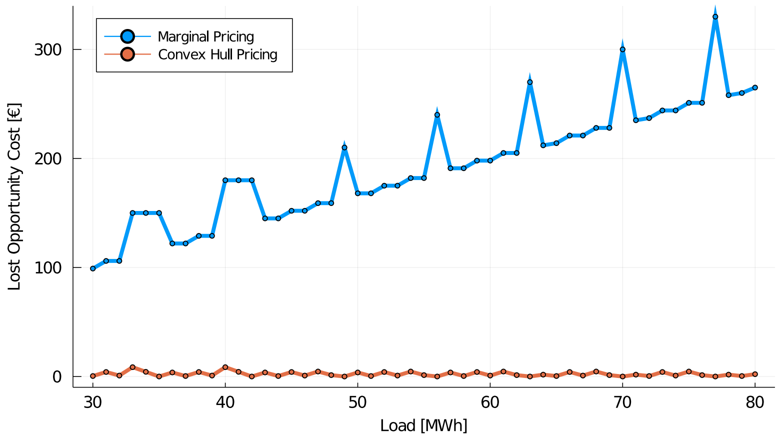


FIGURE 5.3: Lost opportunity costs under discrete investment, for both marginal pricing and convex hull pricing, as a function of the load.

O'Neill et al. (2005) suggest viewing the issue of indivisibilities as one of market incompleteness. One commodity is *energy*, which is sold at the merit order *energy price* (indexed by time and location). Another commodity—that should also be priced—is *capacity* (the discrete investment decisions). In the approach of O'Neill et al. (2005), energy receives a *uniform* price, while capacity is remunerated using *discriminatory payments*. O'Neill et al. (2005) show that there exists a set of prices $(\pi_t^M, \pi_{i,g}^C)$ (remunerating energy and capacities) associated to the allocation $(q_{g,t}^{**}, x_{g,i}^{**})$ (solving problem (5.3)) that is a *competitive equilibrium*. There are two issues with this approach. Firstly, from a practical point of view, it is unclear which actual market mechanism is supposed to output these discriminatory prices[15]. Secondly, from a theoretical point of view, the presumed price-taking behaviour of the suppliers seems to be contradicted by the mere fact that the capacity prices are discriminatory. There is essentially one single supplier for each "investment commodity", and therefore price-taking behaviour seems like wishful thinking[16].

15 The analysis of this chapter has obviously some similarities with the ones of chapter 3, although this chapter studies a long-term problem. The point of the sentence in the text is that, if in the short-term, it is clear how an auctioneer could compute the discriminatory payments based on the bids that are submitted to the auction, it is less clear how such a procedure could be transposed to a long-term context.

16 This puzzling methodological aspect connects to the well-known *Lindahl equilibrium* in public goods (Mas-Colell et al., 1995). The latter also relies on the use of discriminatory prices while assuming price-taking behaviour, and has been subject to the same criticisms from economists.

In this section, we are instead interested in studying the effect of a *uniform* capacity remuneration mechanism (CRM). A capacity market is one form of long-term centralized coordination of investment decisions. The classic arguments in favour of a capacity market rest on its ability to reduce the exercise of market power and its usage as an instrument for hedging investment risk. Instead, this section investigates to what extent it could also turn out to be a means to mitigate the LOC caused by lumpy investments. As in the approach of O'Neill et al. (2005), the set of commodities is extended to include a remuneration for capacity. Nonetheless, the capacity auction that is considered outputs a *uniform* price[17]. Concretely, the profit maximization problem of the market agents is now assumed to be the following:

Definition 5.8 (Supplier Profit Under Energy and Capacity Prices)**.** *The agent g is assumed to maximize its selfish profit function $\mathcal{P}_g$, defined as follows:*

$$\max_{(q,x)_g \in \mathcal{X}_g} \mathcal{P}_g(q, x, \pi^M, \pi^C) \equiv \sum_{t \in \mathcal{T}} \pi_t^M \Delta T_t q_{g,t} - c_g((q,x)_g) + \pi^C \sum_{i \in \mathcal{I}_g} P_{g,i}^{max} x_{g,i} \tag{5.13}$$

The suppliers have two streams of revenue. One comes from selling energy at the marginal energy prices π_t^M under fixed investment. Another comes from the uniform capacity price π^C, which remunerates their installed capacity. The capacity price comes from a capacity auction. Various designs of CRM have been considered in the literature and among practitioners, such as descending clock auctions. Both theory and experience have highlighted the advantages of sealed-bid uniform price auctions (Harbord and Pagnozzi, 2014). Our auction model can be described as follows. The suppliers submit bids that correspond to their investment costs $\sum_i x_{g,i} IC_{g,i}$, discounted by the anticipated short-term surplus from the energy market, $\sum_i P_{g,i}^{max} x_{g,i} \sum_{t \in \mathcal{T}_g} (\pi_t^M - MC_g)$. Here, $\mathcal{T}_g$ are the periods for which the production of plant g is profitable, $\pi_t^M > MC_g$. The system operator is the single buyer for the capacity target C^{min}, which is assumed to be inelastic.

17 A natural extension of the ideas developed in this section would be to consider *several* capacity targets that depend on the technology (as opposed to a *single* aggregated capacity target).

Definition 5.9 (Discrete Capacity Auction). *The capacity auction minimizes the cost of satisfying the inelastic capacity demand C^{min}:*

$$\min_{x} \sum_{g\in\mathcal{G}} \left(\sum_{i\in\mathcal{I}_g} x_{g,i} IC_{g,i} - \sum_{i\in\mathcal{I}_g} P^{max}_{g,i} x_{g,i} \sum_{t\in\mathcal{T}_g} \Delta T_t (\pi^M_t - MC_g) \right) \quad (5.14a)$$

$$(\pi^C) \quad \sum_{g\in\mathcal{G}} \sum_{i\in\mathcal{I}_g} P^{max}_{g,i} x_{g,i} \geq C^{min} \quad (5.14b)$$

$$x_{g,i} \in \{0,1\} \quad \forall g \in \mathcal{G},\ i \in \mathcal{I}_g \quad (5.14c)$$

The literature on CRMs typically focuses on continuous investment settings[18]. This contrasts with how the CRM is implemented in certain countries, such as Belgium, where the auction accepts *only* indivisible bids. In the case of a discrete capacity auction, as in model (5.14), two questions arise: (i) how do we select the bids that are cleared? and (ii) how do we derive the capacity price? Harbord and Pagnozzi (2014) acknowledge these dilemmas in CRMs with indivisibilities. As far as bid selection is concerned, a natural option is to select the cost-minimizing bids, as in model (5.14). Proposition 5.4 establishes the general validity of this approach in continuous settings, while Proposition 5.5 indicates certain limits that are encountered under discrete settings.

Proposition 5.4. *Under a continuous investment model (problem* (5.1)*), with a classical "missing money" problem originating from an energy price cap, there exists a well-calibrated capacity target C^{min} such that the optimal expansion plan x^{**} is also a solution of the capacity auction (i.e. a continuous version of model* (5.14)*).*

Proposition 5.5. *Under a discrete investment model (problem* (5.3)*) with long-term LOC, in some cases, the capacity cleared by the auction (i.e. solving model* (5.14)*) may differ from the optimal expansion plan x^{**} even with a well-calibrated capacity target C^{min}.*

For example, considering Scarf's Example 5.1, for $D = C^{min} = 60$MW, solving the auction of model (5.14) would lead to $x_{Smokestack} = 2$ and $x_{HighTech} = 4$, which corresponds to the optimal mix (cf. Figure 5.2a). On the other hand, solving the same auction for $D = 40$MW would lead to $x_{Smokestack} = 3$ and $x_{HighTech} = 0$, which differs from the optimal mix. This puzzling phenomenon raises the question of how a discrete capacity market should select the bids that are cleared. In practice, alternative

18 This is true for all the previously cited works on CRM (Fabra, 2018; Stoft, 2002; Cramton and Stoft, 2005; Cramton et al., 2013; De Maere d'Aertrycke et al., 2017) that analyze a *continuous* capacity auction.

clearing rules have been used. For instance, according to Elia (2019), the Belgian TSO uses a "heuristic" rule to clear the CRM that even differs from cost-minimization. Moreover, system operators typically perform certain prequalification processes before solving the CRM. In Ontario and certain other systems, the system operator even solves a comprehensive capacity expansion model in order to determine the allocation of the capacity payments (Spees et al., 2013; IESO, 2023). As Proposition 5.5 indicates, this can be justified in certain cases.

As far as the pricing scheme is concerned, Harbord and Pagnozzi (2014) discuss various options, acknowledging the "flexibility in the definition of a market-clearing price" in a *discrete* capacity auction. They essentially focus on alternatives between the highest winning bid and the lowest losing bid. Instead, we will consider that the capacity auction relies on *convex hull pricing* (Definition 5.10). As highlighted in Proposition 5.6, this pricing scheme has the property of mitigating the long-term LOC.

Definition 5.10 (Convex Hull Pricing for Capacity Auctions)**.** *The capacity price π^C is defined as the optimal Lagrangian multiplier*[19] *associated with the market clearing constraint in problem* (5.14)*.*

Proposition 5.6. *The uniform capacity price π^C, as defined in Definition 5.10, minimizes the following lost opportunity costs:*

$$\begin{aligned}\pi^{C*} = \arg\min_{\pi^C \geq 0} \Bigg\{ & \Bigg[\pi^C \Big(\sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{I}_g} P^{max}_{g,i} x^{**}_{g,i} - C^{min} \Big) \Bigg] \\ & + \sum_{g \in \mathcal{G}} \Bigg(\max_{(q,x)_g \in \mathcal{X}_g} \Bigg\{ \sum_{t \in \mathcal{T}} \pi^M_t \Delta T_t q_{g,t} + \pi^C \sum_{i \in \mathcal{I}_g} P^{max}_{g,i} x_{g,i} - c_g((q,x)_g) \Bigg\} \\ & - \Bigg[\sum_{t \in \mathcal{T}} \pi^M_t \Delta T_t q^{**}_{g,t} + \pi^C \sum_{i \in \mathcal{I}_g} P^{max}_{g,i} x^{**}_{g,i} - c_g((q^{**}, x^{**})_g) \Bigg] \Bigg) \Bigg\}\end{aligned} \tag{5.15}$$

Here, (x^{**}, q^{**}) denotes a solution to problem (5.3)*.*

A major question in the capacity auction regards the choice made by the system operator of the capacity target C^{min}. Assuming $C^{min} = \sum_g \sum_i P^{max}_{g,i} x^{**}_{g,i}$ (the optimum of the long term expansion problem (5.3)), then expression (5.15) corresponds to the long-term lost opportunity cost of the suppliers. More generally, as far as the first term (under bracket) in equation (5.15) is concerned, the following result can be established.

19 Considering our simple set of constraints $\mathcal{X}_g$, in our case, computing π^C is equivalent to taking the linear programming relaxation of problem (5.14).

Proposition 5.7. *If $C^{min} \leq \sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{I}_g} P^{max}_{g,i} x^{**}_{g,i}$, then the total LOC of the suppliers under both energy and capacity prices (π^M, π^C) is lower than the LOC under the sole energy price π^M.*

Convex hull pricing in short-term auctions is known to mitigate the *short-term* LOC (Hogan and Ring, 2003). Similarly, Proposition 5.6 shows that CHP in a capacity auction mitigates the *long-term* LOC. However, this positive result has three limits. Firstly, although the capacity price π^C mitigates the LOC, we emphasize that it does not reduce it to zero, thus it does not entirely solve the lumpiness problem. This is, to some extent, expected. While a price cap is a distortion of the energy price that *homogeneously* affects all the suppliers, and may therefore be solved in theory by a uniform capacity price (Cramton et al., 2013), investment indivisibilities distort the energy price in a manner that affects suppliers *heterogeneously*. This implies that it cannot be solved by a single instrument such a uniform capacity price. Secondly, Proposition 5.6 is conditional to the fact that the bids that are cleared in the CRM are coherent with the x^{**}. As highlighted in Proposition 5.5, this may not always be the case. There is no straightforward solution to this problem. In Example 5.6, over the 50 load scenarios, the capacity mix cleared by the CRM does not equal the optimal mix in 11 scenarios (22% of the cases). This also happens in the numerical results of section 5.6, although infrequently. Thirdly—and most importantly—, Proposition 5.6 is also conditional to the right calibration of the capacity targets C^{min}. For instance, as Proposition 5.7 emphasizes, an over-dimensioned capacity target could lead to a capacity price that exacerbates the LOC, as compared to the energy-only market, instead of mitigating it. On the other hand, a capacity target which is too low could drive the CRM price π^C to zero, thereby making the capacity auction pointless. This sensitivity of the success of a capacity auction to the calibration of the capacity target is known. De Maere d'Aertrycke et al. (2017) observe such a sensitivity in a risky environment. We consistently observe it in an environment characterised by the presence of indivisibilities. This sensitivity is revisited in the next section. To sum up, if the results of this section highlight how a CRM *may* partially resolve the incentives to invest in the context of lumpy investments, one has to be careful with the design of the capacity demand curve as well as with the capacity market clearing rule. The following example illustrates the theory that is presented in this section.

Example 5.6. *We consider once again Scarf's Example 5.1. The capacity demand C^{min} is set equal to the optimal capacity mix $\sum_g \sum_i x^{**}_{g,i} P^{max}_{g,i}$. So far, we have considered three settlement schemes: marginal pricing (Definition 5.1), convex*

Settlement Schemes	Energy Price [€/MWh]	Capacity Price [€/MWh]	LOC [€] Smokestack	High Tech	Total
Marginal Pricing	3	/	106	92	198
Convex Hull Pricing	6.2857	/	0.857	0	0.857
Marg. Price + Cap. Price	3	3.2857	0.857	0	0.857

TABLE 5.2: Comparison of the pricing schemes for a demand of 60 MW.

hull pricing (Definition 5.6) and marginal pricing complemented with a uniform capacity price. Table 5.2 presents the prices and LOC results for the three settlement schemes, assuming a market demand of 60 MW. Figure 5.4 reports the lost opportunity costs under these three settlement schemes, for various loads. The red stars in Figure 5.4 flag the load scenarios for which the bids cleared from the capacity auction differ from the optimal solution of the capacity expansion (cf. Proposition 5.5). In these cases, the LOC reported for the CRM assumes that the system operator intervenes for selecting the optimal bids. This could be seen as the most optimistic outcome of a uniform capacity market and is consistent with the separation of primal and dual computations in various short-term auctions, including the EU and US markets. As anticipated from Proposition 5.6, the addition of a capacity payment decreases the total LOC.

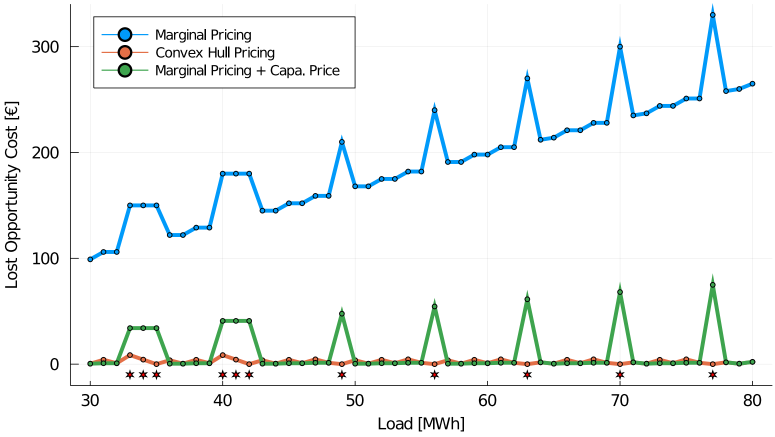


FIGURE 5.4: Lost opportunity costs as a function of the load, for three settlement schemes: (a) marginal pricing (Definition 5.1), (b) convex hull pricing (Definition 5.6) and (c) marginal pricing complemented with an uniform capacity price (Definition 5.10).

5.6 NUMERICAL SIMULATIONS: THE EUROPEAN CAPACITY EXPANSION PROBLEM

We now turn to the quantification of the inefficiencies resulting from the lumpiness of investment in *realistic* settings. We conduct our analysis on an investment model derived from the ENTSO-E capacity expansion model which covers the entire European system.

5.6.1 *The European resource adequacy assessment*

ENTSO-E publishes the European Resource Adequacy Assessment (ERAA) annually. This is an analysis of the adequacy of the pan-European system which assesses European TSOs' ability to ensure security of supply under various scenarios, for a given target year.[20] In the 2021 ERAA study that we consider (ENTSO-E, 2021), the main target year is 2025. The ERAA has two main objectives. The first one is to assess the expected adequacy (measured with the "Loss of Load Expectation" (LOLE) [h/year]), and to compare it to the target LOLE defined by each national TSO for its country. These simulations are performed with *fixed* expected capacity, as foreseen by each national TSO. More related to the current investigation, the second objective is to undertake an Economic Viability Assessment (EVA). This is an adequacy assessment that is based on the capacity mix that results from an *economically viable* investment in power plants. Here, a capacity expansion model is solved, which includes commissioning and decommissioning decisions from the mix that is expected by the national TSOs. In our simulations, we reproduce the model of ERAA (EVA) and use its data to simulate the capacity expansion of the European system[21]. Since the ERAA does not consider integer investment decisions, we slightly adapt the model of ERAA to turn it into a discrete investment model. With this exercise, we are particularly interested in addressing the following questions:

1. How does the introduction of lumpy investment affect the outcome of ERAA? In particular, what would be the magnitude of the LOC in

20 Some more caveats are needed here. It is worth highlighting that "electricity reliability is a function of much more than just adequate investment in generation capacity" (Borenstein et al., 2023). A lot of forced outages are driven not by unreliable power generation assets, but by extreme weather events, transmission or distribution line failures, system operation errors, etc. For instance, a large share of outages in the US "stem from issues related to the delivery, rather than the production, of electricity" (Borenstein et al., 2023). In this chapter, as in ERAA, we neglect these important drivers of system reliability (which would tend to qualify the usage of capacity markets), and we focus instead on the sole issue of resource adequacy.

21 The data can be retrieved from the website of ERAA.

such a large discrete investment model, that includes many technologies and nodes? This aims at illustrating numerically the importance of lumpiness of investment advocated in sections 5.3 and 5.4.

2. How would a discrete CRM affect the incentives of agents to invest in such a realistic case study? This aims at illustrating numerically the theory of section 5.5.

We notice that ERAA also includes an analysis of the impact of a CRM. However, our analysis fundamentally differs from ERAA. Under the *continuous* setting considered by ENTSO-E, the CRM is used to solve the *missing money* problem. Indeed, the study of ENTSO-E shows that, when running the capacity expansion model (EVA) without a capacity market but with a price cap in the energy market at 15k€/MWh, the new mix of capacities results in a slight under-investment. Concretely, the energy market alone does not lead to the "optimal" investment, as defined by the LOLE targets. In this continuous case, the capacity market is needed because of the *flawed price cap*, which is not consistent with the LOLE target[22]. In our *discrete* case, we assume that the price cap of 15k€/MWh reflects the right VOLL, such that there is no classical "missing money". Instead, as we work with discrete investments, the CRM plays a role of mitigating the long-term lost opportunity cost.

5.6.2 *The ERAA model*

The detailed mathematical model of EVA is provided in appendix 5.B. In a nutshell, the EVA model includes 37 countries modelled as 59 bidding zones. The power grid is composed of HVAC and HVDC lines, although in the EVA the network constraints are represented using an ATC model. The model considers various climate years, that can be viewed as a set of 31

22 The price cap is a key driver of the investment decisions in an energy-only market. It reflects the value at which the "lost load" is priced in the energy market (the VOLL). It should be aligned with the LOLE targets in order for the market to induce the right level of investment. Indeed, there is a strong connection between the VOLL and the LOLE (Stoft, 2002; Cramton and Stoft, 2005). Concretely, the invested capacity will be optimal if it is such that the marginal cost of an additional MW of peaking capacity (the investment cost of a peaker, IC_{peak}) equals the cost of one more MW of blackout ($VOLL \times LOLE$). This can be summarized by the following equation: $IC_{peak} = VOLL \times LOLE$. Note that this relationship can be derived from the KKT conditions (5.2) of the continuous investment problem (5.1), with $V_t = VOLL$. Indeed, considering the peaking units, we deduce from equation (5.2b) that $IC_{peak} = \sum_{t \in \mathcal{T}} \Delta T_t \mu_{peak,t}$. Either the peaker sets the price and $\mu_{peak,t} = 0$, or demand sets the price. In the latter case, neglecting the unlikely situation in which there is a price indeterminacy, the energy price will soar up to the $VOLL$, so that $\mu_{peak,t} = VOLL - MC_{peak} \approx VOLL$. From which we conclude that $IC_g \approx VOLL \times LOLE$. See also Papavasiliou (2024).

scenarios[23] of load and renewable production. Not serving the load (load curtailment) is priced at $VOLL$. Production curtailment is not penalized in the objective function (the model assumes *free disposal*). The operational constraints are convex, and so are the investment decisions, which are all continuous. All the power plants of the same technology in a bidding zone are aggregated into one large virtual power plant. There are six main types of generation assets. (i) *Existing* plants can be partially retired, leading to a fixed cost reduction (such as yearly maintenance costs). (ii) *New* plants can be constructed with a fixed cost. These are the two investment decisions: continuous variables x_g^{new} (commissioning) that come with an investment cost IC_g^{new} and x_g^{exist} (decommissioning) that save an investment cost IC_g^{exist}. (iii) Renewable assets are exogenous and therefore directly integrated in the net load. (iv) Demand response, essentially an *elastic load* (or "load shedding"), is modelled as an additional convex generator at a given price. (v) Batteries are essentially a *load shifting* asset, and are modelled as a unique battery per node. (vi) There are four different types of hydro plants (all convex, described in appendix 5.B).

Figure 5.5 provides an overview of the merit order of the entire ENTSO-E system (i.e. the operational cost MC_g). As far as the investment decisions are concerned, each of the technologies of Figure 5.5 could be decommissioned, while the commissioning decisions are limited to two technologies, CCGT and OCGT plants. To provide an order of magnitude, their investment costs are respectively 143,000 and 95,000 €/MW/y. The decommissioning of generation assets of technology g is limited by a parameter $RCapa_g^{max}$ that is provided by ENTSO-E. This parameter is either set to the installed capacity (meaning that the technology could be entirely decommissioned) or to a lower limit in case ENTSO-E considers it unrealistic to decommission entirely the technology (e.g. nuclear plants in France are not allowed to be decommissioned). The commissioning of new OCGT and CCGT plants is limited by a parameter $Capa_g^{max}$.

For our experiments, we have modified the ERAA model in two ways (the detailed models are in appendix 5.B):

- Since we are interested in the discrete investment model, the continuous investment decisions of ERAA are converted to integers. Concretely, investments are now the variables $x_g^{new}, x_g^{exist} \in \mathbb{N}$ which stand for investments in integer numbers of capacity lumps, modelled by parameters C_g^{new} and C_g^{exist}, which are technology specific (for example, a CCGT unit is 500 MW, an OCGT unit is 300 MW... e.g. $x_{CCGT}^{new} = 3$ means the entrance of 3 CCGT units of 500 MW each).

23 ENTSO-E provides 35 scenarios (climate years). As we were not computationally able to solve 4 of them (the climate years 1988, 2000, 2005, 2006), we only use 31 scenarios in our analysis.

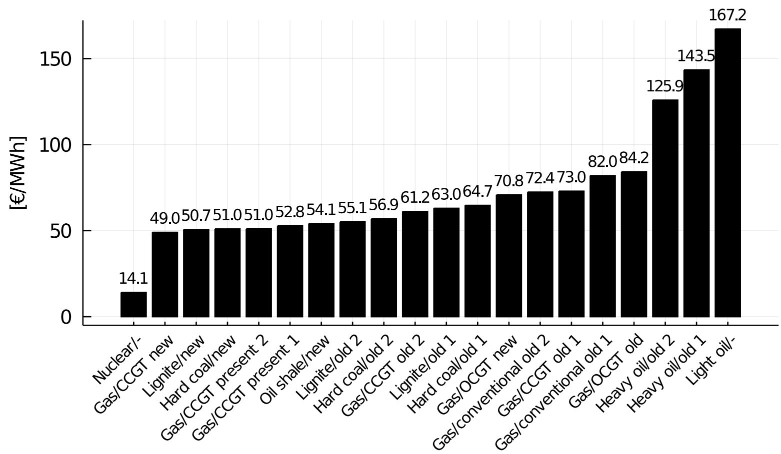


Figure 5.5: Merit order of the EVA model for the entire ENTSO-E region. Note that the model also includes a carbon tax of 40€ per ton of CO_2 which is directly included in the operating cost of the plants.

The comprehensive data for parameters C_g^{new}, C_g^{exist} is provided in Table 5.8 of appendix 5.B. The energy prices $\pi_{i,t}^M$ in this model are assumed to be the merit order prices of Definition 5.1.

- In the same spirit as section 5.5, a capacity market is introduced. The capacity market is assumed to remunerate the capacity of flexible generation units only (x_g^{new}, x_g^{exist}), i.e. the capacity auction is limited to the thermal units (DSR, renewable or hydro plants cannot participate). As compared to section 5.5, the capacity targets C_i^{min} defined by the system operator are now indexed by the bidding zone i.

The model is implemented in Julia (JuMP) and is solved with Gurobi. The computations are performed on the Lemaitre3 cluster (80 nodes with two 12-core Intel SkyLake 5118 processors at 2.3 GHz and 95 GB of RAM), which is hosted at the Consortium des Equipements de Calcul Intensif (CECI).

5.6.3 *Numerical results*

We simulate three models: the continuous "vanilla" version of ERAA, the discrete version, and the latter complemented by capacity payments.

	Total Cost			Commissioning		Decommissioning		LOC
Scenarios	Cont.	Disc.	Inc.	Cont.	Disc.	Cont.	Disc.	Disc.
...								
2025/7	7.385e10	7.409e10	0.3%	4745	3800	37790	33000	4.91e8
...								
2025/29	7.228e10	7.258e10	0.4%	3690	3300	46929	43100	4.254e8
...								
Average	7.614e10	7.634e10	0.3%	7554	7048	29560	25920	1.139e9

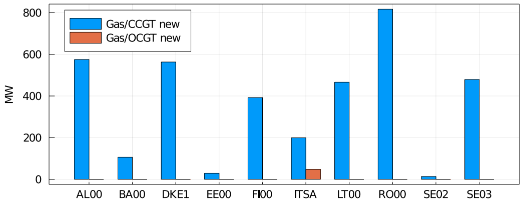
TABLE 5.3: Comparison of the discrete and continuous results of ERAA (the full results are in appendix 5.C).

The simulations are performed over 31 scenarios[24] (historical load and climate years projected to 2025 market conditions). Tables 5.3 and 5.4 report the average results of 31 scenarios as well as the detailed results for two scenarios, 2025/7 and 2025/29. The full results are in appendix 5.C. We highlight three main sets of observations relative to the comparison of discrete versus continuous investment settings, the magnitude of the long-term lost opportunity cost, and the effect of a CRM.

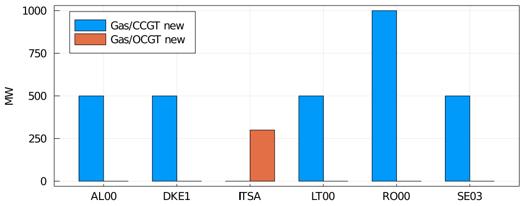
Firstly, as far as the comparison of the discrete and continuous model is concerned, Table 5.3 summarizes the main results from the simulations. We observe that both models can lead to fairly different results of commissioning and decommissioning decisions. Figure 5.6a illustrates these differences on scenario 2025/29. We observe that the commissioning of new capacities output by the continuous version of the model is reallocated across the bidding zones because of the lumpiness of the capacity. More importantly, Table 5.3 also reports the total cost under both continuous and discrete models. We observe that the total costs are strikingly similar. The lumpiness of investment decisions marginally affects the total system costs, which increase by a mere 0.3% on average.

Secondly, if the lumpiness of investment has a minor effect on *costs*, it can however significantly affect the *incentives* of the market agents. Indeed, in the continuous case, the lost opportunity cost of *all* the new and existing units is zero. This is anticipated from the theory. Under a convex model, the uniform energy prices together with the allocation of resources (investment and dispatch) form a competitive equilibrium. Instead, in the discrete case, the market agents are not in equilibrium. This is quantified in Table 5.3: on

24 This is consistent with the methodology of ENTSO-E (2021): ENTSO-E performs the simulations separately on multiple years (scenarios) and then averages the expansion plans.



(a) Continuous model



(b) Discrete model

FIGURE 5.6: Commissioning decisions under the continuous and discrete model for scenario 2025/29.

average, the total LOC stands for 1.5% of the total system cost. Both the new and existing power plants face incentives to deviate from the welfare maximizing allocation. We further focus on scenario 2025/29. Among all the possible comissioning (resp. decomissioning) decisions, 11% (resp. 10%) face a positive LOC. These figures show that the LOC is not contained to a few plants at the margin, but it affects the investors more broadly. At the same time, this LOC—the "burden" of investments' indivisibilities—is not split uniformly over the entire system, but it rests on the shoulder of some private investors. For example, the revenue shortfall faced by the OCGT plant installed in ITSA (Table 5.5) stands for 63% of its investment cost. More generally, 67% of the effective commissioning decisions come with a revenue shortfall. On average, this revenue shortfall corresponds to 22% of the investment cost.

The lost opportunity costs are further decomposed into revenue shortfall and foregone opportunities in Table 5.4. A *revenue shortfall* should be read as follows. For a *new* plant, it means that it is asked to be constructed while not covering its investment cost. For an *existing* plant, it means that it is asked to not be decommissioned despite facing damages. This is further illustrated in Tables 5.5 and 5.6 which report a sample of the financial standing of various technologies per bidding zone for scenario 2025/29. In Table 5.5, we observe various new plants that are commissioned (the CCGT units in zones DKE1, AL00 and RO00 as well as the OCGT in zone ITSA) while suffering losses. As far as the existing plants are concerned, in Table

Scenarios		Without capacity market: New units	Exist units	Total	With capacity market: Inelastic	Elastic	No Coord.
...							
	LOC	3.534e8	1.376e8	4.91e8	4.863e8	6.354e8	1.144e9
2025/7	*RS*	1.802e7	0.0	1.802e7	1.335e7	4.154e7	3.244e7
	FO	3.354e8	1.376e8	4.73e8	4.73e8	5.939e8	1.111e9
...							
	LOC	1.052e8	3.202e8	4.254e8	5.447e7	2.678e8	1.328e9
2025/29	*RS*	6.925e7	3.171e8	3.864e8	1.312e7	4.362e7	3.012e7
	FO	3.599e7	3.061e6	3.905e7	4.135e7	2.242e8	1.297e9
...							
	LOC	6.7e8	4.692e8	1.139e9	7.192e8	9.044e8	1.429e9
Average	*RS*	9.805e7	3.376e8	4.356e8	1.477e7	2.364e7	2.429e7
	FO	5.72e8	1.316e8	7.037e8	7.045e8	8.808e8	1.405e9

TABLE 5.4: Analysis of investor incentives decomposed into lost opportunity costs (*LOC*), revenue shortfall (*RS*) and foregone opportunity (*FO*), for the two cases including or not a capacity payment. The results report three CRM settings: the inelastic capacity target, the elastic capacity demand curve and the inelastic capacity target computed without European coordination (the full results are in appendix 5.C).

5.6 we observe many technologies (the table only shows a sample) that are required to stay in the market while they suffer losses (e.g. some oil plants in zone GR03 as well as CCGT units in FR00, BE00, HU00 or PT00, or some lignite plants in RS00). Similarly, twelve CCGT units in ES00 leave the market while even more units would prefer to leave the market due to the fact that they are not profitable.

A *foregone opportunity* should be understood as follows. For a *new* technology, it means that there is an incentive to invest more than what is socially optimal. Some technologies are not investing at all, while they would be profitable. Others are investing, but less than what they would given the energy price signal. For example, in Table 5.5, we observe that no CCGT plants in zone SE04 are commissioned while they would be profitable. In zones LT00 and SE03, one CCGT is commissioned while it is profitable and would therefore have incentives to expand. The last case means that certain new CCGT plants not only have incentives to deviate from the welfare-maximising allocation but they also earn a non-zero profit for a resource that is not scarce. They earn a "discreteness rent" of 3740

Zone	Technology	Investment	Profit	*LOC*	*RS*	*FO*
SE04	CCGT new	0×500	0.0	3.05e6	0.0	3.05e6
DKE1	CCGT new	1×500	$-1.871e6$	1.871e6	1.871e6	0.0
LT00	CCGT new	1×500	2.007e6	2.007e6	0.0	2.007e6
AL00	CCGT new	1×500	$-1.626e7$	1.626e7	1.626e7	0.0
FI00	CCGT new	0×500	0.0	1.164e7	0.0	1.164e7
EE00	CCGT new	0×500	0.0	7.328e6	0.0	7.328e6
RO00	CCGT new	2×500	$-3.284e7$	3.284e7	3.284e7	0.0
SE02	CCGT new	0×500	0.0	3.468e6	0.0	3.468e6
SE01	CCGT new	0×500	0.0	2.199e6	0.0	2.199e6
SE03	CCGT new	1×500	1.733e6	3.466e6	0.0	3.466e6
ITSA	OCGT new	1×300	$-1.828e7$	1.828e7	1.828e7	0.0
LV00	CCGT new	0×500	0.0	2.829e6	0.0	2.829e6

Table 5.5: Detailed analysis of agent incentives for the *new* plants (commissioning) for scenario 2025/29.

€/MW/year (for a 500MW CCGT it means 1.87 M€/year). For an *existing* plant, a foregone opportunity means that it is asked to retire while the plant is profitable. In Table 5.6, several CCGT plants in zone UK00 are asked to retire while they are profitable.

These results confirm the theoretical findings from section 5.4: in an investment problem, the LOC resulting from indivisibilities can be significant, even in large systems. Certain market agents face incentives to invest more than what is socially optimal. In practice, they may not invest but they will then collect a positive rent for a resource that is not scarce. Other agents cannot cover both their operational and capital costs. The energy price does not play well the coordination role that it fulfils in convex settings, nor does it convey the information properly. Indeed, in the *discrete* case, some technologies end up with positive (or negative) profits. But, as Scarf emphasises, and unlike what would happen in the *continuous* case, the fact that a technology faces a positive (negative) profit does not indicate that the entire system welfare could be improved by increasing (decreasing) the investment in that technology—in fact, it would not. There is no easy solution to this issue, and as we observe later, the introduction of a capacity market can make matters worse if not properly calibrated.

The third and last aspect of our analysis regards the impact of a uniform capacity remuneration. We test three shapes of capacity demand curve:

Zone	Technology	In Place	Investment	Profit	*LOC*	*RS*	*FO*
DKE1	Light oil	529	-2×100	-2.871e6	2.618e6	2.618e6	0.0
GR03	Light oil	277	0×100	-3.632e6	2.622e6	2.622e6	0.0
PT00	CCGT present 1	990	0×450	-1.367e6	1.243e6	1.243e6	0.0
HU00	CCGT old 2	976	0×400	-1.726e7	6.407e6	6.407e6	0.0
BE00	CCGT present 2	3550	0×450	-6.709e6	5.953e6	5.953e6	0.0
FR00	CCGT present 2	5148	0×450	-7.521e7	7.231e7	7.231e7	0.0
UK00	CCGT old 2	15010	-3×400	3.173e7	2.757e6	0.0	2.757e6
UK00	CCGT old 1	593	-1×400	146800.0	303600.0	0.0	303600.0
RS00	Lignite new	1106	0×300	-2.863e7	2.33e7	2.33e7	0.0
ES00	CCGT present 1	24499	-12×450	-1.492e8	5.452e7	5.452e7	0.0

TABLE 5.6: Detailed analysis of agent incentives for *existing* plants (decommissioning) for scenario 2025/29 (sample).

(A) An *inelastic* capacity demand with the national capacity targets C_i^{min} set to the *optimal* investment target with European coordination (i.e. solving the expansion problem), as in Proposition 5.6: $C_i^{min} = \sum_{g \in \mathcal{G}_i^{new}} x_g^{new**} C_g^{new} - \sum_{g \in \mathcal{G}_i^{exist}} x_g^{exist**} C_g^{exist}$, where x_g^{new**} and $x_g^{exist**}$ are the optimal investment decisions derived from solving the discrete investment problem. An example is provided in Figure 5.7.

(B) An *elastic* capacity demand which follows the design proposals in the literature (Cramton and Stoft, 2005; Cramton et al., 2013) as well as practical applications (see the survey of Papavasiliou (2021)). An illustration is provided in Figure 5.7. The demand for capacity is worth two times the entry cost of a peaker (here, an OCGT unit) up to C_i^{min} (the *optimal* investment target) minus 5%. Then the valuation for capacity decreases sharply to one times the entry cost of a peaker at C_i^{min}, and finally becomes zero at C_i^{min} plus 15%.

(C) An inelastic capacity demand, but with the C_i^{min} targets computed *without* European coordination. In this case, each country computes the target capacity independently, instead of solving the European investment problem. Concretely, in order to compute C_i^{min}, we simulate an adapted version of the capacity expansion problem of ENTSO-E, where each country i is treated as an island, having to meet its national load only with domestic capacity. An illustration is provided in Figure 5.7.

The right half of Table 5.4 presents how the lost opportunity costs are affected by the addition of a capacity market. Under CRM model (A),

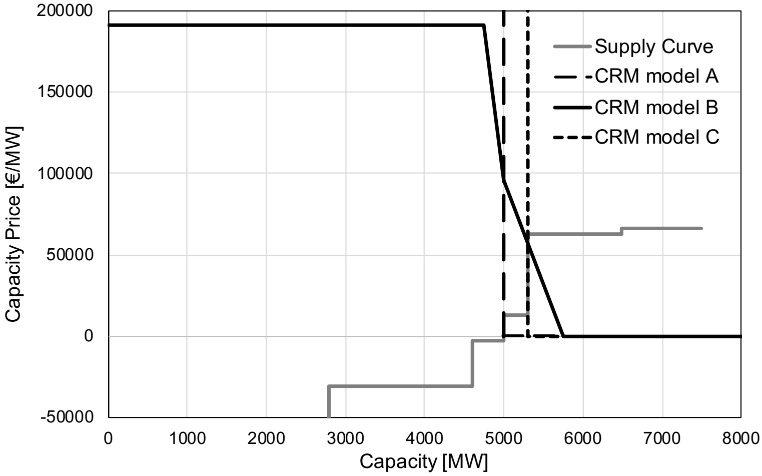


FIGURE 5.7: Illustration of the capacity demand curve for scenario 2025/29 in Ireland for models (A) (inelastic demand), (B) (elastic demand) and (C) (no EU-coordination). In these models, the Irish capacity price is, respectively, 0, ∼60,000 and ∼13,000 €/MW/y. As a point of comparison, the capacity price of the Belgian CRM in 2022 was ∼50,000 €/MW/y.

and as expected from Proposition 5.6, the capacity payments improve the overall incentives of the market agents. On average, the long-term lost opportunity costs decrease by 40% following the inclusion of a CRM, while the revenue shortfalls drop by 97%. Nonetheless, the magnitude of the impact of the CRM is heterogeneous across scenarios: for example, if the effect is significant in scenario 2025/29, it is less so in scenario 2025/7. We notice that, in our computations, the capacity price remunerates the optimal capacity mix. This could be regarded as the most optimistic result that can be achieved by a uniform price CRM. Indeed, as highlighted in Proposition 5.5, it may happen that the bids cleared by the CRM (as problem (5.14)) differ from the capacity mix optimizing problem (5.3). For example, in scenario 2025/29, 20 zones out of the 59 have a positive capacity price. Among these 20 zones, discrepancies between the CRM results and the optimal expansion plan occur in 15% of the cases.

The two other CRM designs stand for plausible cases of an over dimensioned target C_i^{min}. They aim at evaluating the impact of a capacity price as soon as the capacity demand curve departs from the idealized settings of Proposition 5.6. As indicated by Proposition 5.7, an over-dimensioned capacity target can exacerbate the LOC. Under CRM model (B), we observe that, on average, the addition of a capacity payment still improves the incentives of market agents compared to the sole energy remuneration. In scenario 2025/29, the CRM model (A) allows to cut by ten the total lost opportunity costs. Model (B) does not perform as well, nevertheless it still lowers by 40% the total LOC compared to an energy-only settlement. However, scenario 2025/7 also highlights how model (B) can not only fail to achieve the same performance as model (A), but also perform worse than an energy-only market. In this case, the addition of a capacity payment makes the lost opportunity costs worse than they are under the sole marginal energy price. As shown in Table 5.4, CRM model (C) has a more disruptive effect on agents' incentives. Since it neglects international coordination, this model tends to increase the capacity demanded in each country, thereby amplifying foregone opportunities. This highlights the benefits of having a European coordination in defining the national CRM targets, in the spirit of ERAA.

5.7 CONCLUSION

In this chapter, we analyse the problem of indivisibilities in investment decisions and their impact on the ability of a decentralized energy market to support efficient investments. We analyse the market failure that occurs under indivisible investment. This failure can be measured by the concept of *long-term* lost opportunity cost, which is introduced in the chapter. This lost opportunity cost prevents a purely decentralized energy market to lead to a long-term equilibrium. Indivisibilities in investment have often been overlooked in the literature. A persistent argument for neglecting indivisibilities is that they supposedly vanish when the market size increases. We accurately reconstruct the underlying theoretical argument, by reviewing a classical result from the theory of general equilibrium, that we transpose to the context of power markets. We highlight that this result is only valid under specific pricing assumptions. We show that, as far as the investment problem is concerned, under the classic "merit order pricing", the long-term lost opportunity costs can be arbitrary large. This theoretical argument is confirmed by our numerical simulations.

In order to address this market failure, we analyse the effect of introducing a CRM. We show that investment indivisibilities cast a new light on the role played by a CRM. We particularly propose the novel concept of convex hull pricing (CHP) for capacity auctions. We show that, similarly to CHP in short-term auctions, it can mitigate *long-term* lost opportunity costs. Nevertheless, we also stress the limits of a CRM: we highlight that its effect can be inconclusive—and even counter-productive—when the

CRM is ill-designed. We illustrate these findings on the realistic capacity expansion model used by ENTSO-E for assessing the capacity adequacy of the European system.

As future work, we envision three possible directions. From a theoretical perspective, this work treats indivisibilities in isolation from other imperfections such as market power or risk. One theoretical inquiry is to what extent these imperfections, when combined, reinforce or mitigate each other. From a computational perspective, we have introduced indivisibilities in the capacity expansion model of ENTSO-E. This model has been work in progress for several years. A recent upgrade is the introduction of uncertainty in the model (Ávila et al., 2023). Future work could focus on combining the two features (uncertainty and indivisibility) in one model. Finally, from a policy perspective, our work focused on the interplay between investment indivisibilities and capacity markets. We have highlighted two main problems that could be explored in future works. One further development could attempt to find bounds on the capacity demand C^{min}, as safeguards against over-dimensioning. Another development could explore whether some heuristics could guide the capacity market towards the optimal mix x^{**}, without having to rely on a comprehensive capacity expansion model.

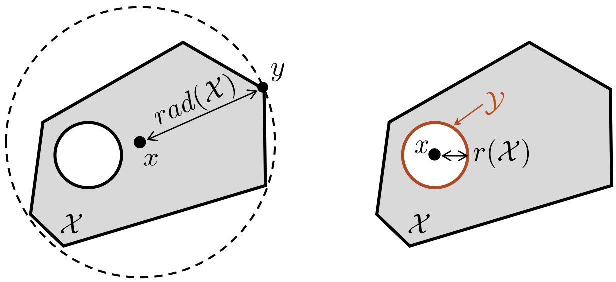


FIGURE 5.8: Illustration of the concepts of *radius* (left figure) and *inner radius* (right figure) of a non-convex set $\mathcal{X}$ (here, $\mathcal{X}$ is a polytope with a hole).

APPENDICES

5.A PROOFS OF THE PROPOSITIONS

Before establishing the proof of Proposition 5.1, we recall the concept of the *inner radius* of a non-convex set.

Definition 5.11. *Let $\mathcal{X} \subset \mathbb{R}^n$ be a compact non-convex set. The radius of this set is the radius of the smallest ball containing the set:*

$$rad(\mathcal{X}) = \min_{x \in \mathbb{R}^n} \max_{y \in \mathcal{X}} |x - y|$$

For any $x \in \text{conv}(\mathcal{X})$, there is a set $\mathcal{Y}$ spanning x (i.e. $x = \sum_{y \in \mathcal{Y}} \lambda(y) y$, with $\lambda(y) \geq 0$ and $\sum_{y \in \mathcal{Y}} \lambda(y) = 1$). The inner radius of $\mathcal{X}$ is the radius of the smallest ball including the smallest set $\mathcal{Y}$ spanning x, for any $x \in \text{conv}(\mathcal{X})$:

$$r(\mathcal{X}) = \max_{x \in \text{conv}(\mathcal{X})} \min_{\substack{\mathcal{Y} \subset \mathcal{X} \\ \text{spans } x}} rad(\mathcal{Y})$$

Both concepts are illustrated in Figure 5.8. Note that, if the production set is convex, its inner radius is clearly 0.

Proof of Proposition 5.1. Let π^{CH} denote the convex hull prices and (q^*, x^*, d^*) the associated allocation in the relaxed problem, i.e. $(q^*, x^*)_g \in \text{conv}(\mathcal{X}_g)\ \forall g$. Since we assume that the consumption set $\mathcal{X}_d$ is convex, we can set $d^\dagger = d^* \in \mathcal{X}_d$, which indeed solves problem (5.6) under price π^{CH}.

Regarding the production plan, by definition of the convex hull, $(q^*, x^*)_g = \sum_i \alpha_{g,i} (q^\dagger, x^\dagger)_{g,i}$ with $\sum_i \alpha_{g,i} = 1$, $\alpha_{g,i} > 0$, for some $(q^\dagger, x^\dagger)_{g,i} \in \mathcal{X}_g$. We denote $\mathcal{Y}_g \subset \mathcal{X}_g$ as the smallest set of those $(q^\dagger, x^\dagger)_{g,i}$. Clearly, all the points

$(q^\dagger, x^\dagger)_{g,i} \in \mathcal{Y}_g$ solve problems (5.5) under the price π^{CH} (by optimality of the allocation (q^*, x^*) in the relaxed problem).

We define $\widehat{\mathcal{Y}_g} = \{q | (q,x) \in \mathcal{Y}_g\}$ (the projections of the previously defined sets $\mathcal{Y}_g$ over the space of variables q), and $z^* \in \mathbb{R}^{|\mathcal{T}|} : z^*_t = \sum_g q^*_{g,t} \in \sum_g \text{conv}(\widehat{\mathcal{Y}_g})$. By Starr's Theorem[25], there exists a $z^\dagger \in \sum_g \widehat{\mathcal{Y}_g}$ such that $|z^* - z^\dagger| \leq \sqrt{|\mathcal{T}|}A$. □

Proof of Proposition 5.2. The proof proceeds in two steps. Firstly, from the central result of convex hull pricing theory (Gribik et al., 2007), the total lost opportunity cost corresponds to the *duality gap*:

$$\sum_{g \in \mathcal{G}} LOC_g(\pi^{CH}) + LOC_d(\pi^{CH}) = z^*_D - z^*_P$$

To see this, it suffices to write the Lagrangian relaxation corresponding to z^*_D and to rearrange the terms. Secondly, this duality gap is bounded[26]. Indeed, using Minkowski's extended formulation, problem (5.8) can be written as the following *linear* program:

$$z^*_D = \max_{d_t, \lambda^k_g} \sum_{t \in \mathcal{T}} \Delta T_t d_t V_t - \sum_{g \in \mathcal{G}} \sum_{k \in K_g} \lambda^k_g \hat{c}^k_g \tag{5.16a}$$

$$\sum_{g \in \mathcal{G}} \sum_{k \in K_g} \lambda^k_g \hat{q}^k_{g,t} \geq d_t \qquad \forall t \in \mathcal{T} \tag{5.16b}$$

$$0 \leq d_t \leq D_t \qquad \forall t \in \mathcal{T} \tag{5.16c}$$

$$\sum_{k \in K_g} \lambda^k_g = 1 \qquad \forall g \in \mathcal{G} \tag{5.16d}$$

$$\lambda^k_g \geq 0 \qquad \forall g \in \mathcal{G},\ k \in K_g \tag{5.16e}$$

The set K_g denotes the number of extreme points of $\mathcal{X}_g$ (which is assumed to be a *compact* set). Parameters $\hat{q}^k_{g,t}$ and $\hat{c}^k_g$ denote the production schedule and cost associated to each extreme point k of $\mathcal{X}_g$. There are $|\mathcal{T}|$ variables d_t and $|K_g|$ variables λ^k_g for each of the $|\mathcal{G}|$ suppliers. From constraint (5.16d), there is *at least* one non-zero λ^k_g per supplier g. If there is *exactly one*, the

25 (Starr (1969), a corollary of the Shapley-Folkman Theorem) Let $\mathcal{X}_i \subset \mathbb{R}^n$ be m non-convex sets such that $r(\mathcal{X}_i) \leq A\ \forall i = 1...m$, and let $x \in \text{conv}(\sum_{i=1...m} \mathcal{X}_i) \subset \mathbb{R}^n$. Then, there exists a $y \in \sum_{i=1...m} \mathcal{X}_i$ such that $|x - y| \leq \sqrt{n}A$.

26 There are two ways to derive this bound. One immediately relies on the Shapley-Folkman theorem (which is also the crux of the Arrow-Starr proofs). See, for instance, Bertsekas and Sandell (1982) for the application of the Shapley-Folkman theorem to the estimation of the duality gap of a separable non-convex optimization problem. Another way relies on basic linear programming theory, that we use here. Our reasoning is adapted from Bertsekas et al. (1983).

solution of the relaxed problem is feasible. From the fundamental theorem of linear programming theory, there is an optimal solution that has *at least* as many constraints as variables that are tight. Therefore, we know there are *at most* $|\mathcal{T}|$ additional non-zero λ_g^k, meaning that at most $|\mathcal{T}|$ suppliers have more than one $\lambda_g^k > 0$ (a production plan that is *infeasible*). Starting from the production plan that solves the convex relaxation, *at most* $|\mathcal{T}|$ production plans should be modified in order to obtain a *feasible* primal solution (under-approximating z_P^*). This modification costs *at most* ρ_g, which is the maximum cost resulting from turning an infeasible production plan $(\hat{q}, \hat{x})_g \in \text{conv}(\mathcal{X}_g)$ to a feasible production plan that supplies at least as much power. We conclude that $z_D^* - z_P^* \leq \rho|\mathcal{T}|$. □

From the proof, it is obvious that an alternative bound (tighter in case the ρ_g vary significantly between the power units) is $z_D^* - z_P^* \leq \sum_{g \in \mathcal{G}^{max}} \rho_g$, with $\mathcal{G}^{max}$ being the set of $|\mathcal{T}|$ generators with the highest ρ_g. Furthermore, for a convex production set $\mathcal{X}_g$, clearly $\rho_g = 0$.

Proof of Proposition 5.3. The Proposition follows immediately from Example 5.4. □

The next proof is adapted from Theorem 1 in Papavasiliou (2021).

Proof of Proposition 5.4. In convex settings, the classical "missing money", which motivates the use of a capacity market, arises because of a price cap in the energy market. Let us assume that $\pi_t^M = PC$ when the system is short (load is curtailed). The continuous capacity market is:

$$\min_{q,x} \sum_{g \in \mathcal{G}} \left(x_g IC_g - \sum_{t \in \mathcal{T}} \Delta T_t (\pi_t^M - MC_g) q_{g,t} \right) \tag{5.17a}$$

$$(\pi^C) \quad \sum_{g \in \mathcal{G}} x_g \geq C^{min} \tag{5.17b}$$

$$0 \leq q_{g,t} \leq x_g \qquad \forall g \in \mathcal{G},\ t \in \mathcal{T} \tag{5.17c}$$

$$x_g \geq 0 \qquad \forall g \in \mathcal{G} \tag{5.17d}$$

We show that there is a well-calibrated C^{min} such that the optimal solution x^* of problem (5.1) is also a solution of auction (5.17). The KKT conditions of problem (5.17) are:

$$0 \leq q_{g,t} \perp MC_g - \pi_t^M + \mu_{g,t} \geq 0 \quad \forall g \in \mathcal{G},\ t \in \mathcal{T} \tag{5.18a}$$

$$0 \leq x_g \perp IC_g - \pi^C - \sum_{t \in \mathcal{T}} \Delta T_t \mu_{g,t} \geq 0 \quad \forall g \in \mathcal{G} \tag{5.18b}$$

$$0 \leq x_g - q_{g,t} \perp \mu_{g,t} \geq 0 \quad \forall g \in \mathcal{G},\ t \in \mathcal{T} \tag{5.18c}$$

$$0 \leq \sum_{g \in \mathcal{G}} x_g - C^{min} \;\perp\; \pi^C \geq 0 \tag{5.18d}$$

Let x^* be the solution of problem (5.1). We want to show that it satisfies (5.18). The energy price only differs between problems (5.1) (where it is called π_t) and (5.18) (π_t^M) during the scarcity periods $\mathcal{ST}$, i.e. for $t \in \mathcal{ST}$, $\pi_t^M = PC$ while $\pi_t = V_t$. We denote by $\mu_{g,t}^*$ the scarcity rents μ solving (5.2). Outside the scarcity periods, equations (5.18a) are equivalent to (5.2a) and $\mu_{g,t} = \mu_{g,t}^*$. During the scarcity periods $\mu_{g,t}^* = V_t - MC_g$ while equations (5.18a) can be written as $\mu_{g,t} = PC - MC_g = \mu_{g,t}^* - V_t + PC$. Equations (5.2b) can then be written equivalently as:

$$\begin{aligned} 0 \leq x_g \;\perp\; IC_g - \pi^C - \sum_{t \in \mathcal{T}} \Delta T_t \mu_{g,t}^* & \\ + \sum_{t \in \mathcal{TS}} \Delta T_t (V_t - PC) \geq 0 \quad \forall g \in \mathcal{G} & \end{aligned} \tag{5.19}$$

Defining $C^{min} = \sum_{g \in \mathcal{G}} x_g^*$, equation (5.18d) implies $\pi^C \geq 0$. Fixing $\pi^C = \sum_{t \in \mathcal{TS}} \Delta T_t (V_t - PC)$, equation (5.19) is then equivalent to (5.2b). □

Proof of Proposition 5.5. The Proposition follows immediately from Example 5.6. □

Proof of Proposition 5.6. Capacity market (5.14) can be written equivalently as follows:

$$\min_{q,x} \sum_{g \in \mathcal{G}} \left(\sum_{i \in \mathcal{I}_g} x_{g,i} IC_{g,i} - \sum_{t \in \mathcal{T}} \Delta T_t (\pi_t^M - MC_g) q_{g,t} \right) \tag{5.20a}$$

$$(\pi^C) \quad \sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{I}_g} P_{g,i}^{max} x_{g,i} \geq C^{min} \tag{5.20b}$$

$$0 \leq q_{g,t} \leq \sum_{i \in \mathcal{I}_g} P_{g,i}^{max} x_{g,i} \quad \forall g \in \mathcal{G},\ t \in \mathcal{T} \tag{5.20c}$$

$$x_{g,i} \in \{0, 1\} \quad \forall g \in \mathcal{G},\ i \in \mathcal{I}_g \tag{5.20d}$$

The Lagrangian dual problem is then (rearranging the terms):

$$\min_{\pi^C \geq 0} \left\{ -C^{min} \pi^C + \sum_{g \in \mathcal{G}} \max_{(q,x)_g \in \mathcal{X}_g} \left\{ \sum_{t \in \mathcal{T}} \Delta T_t \pi_t^M q_{g,t} + \pi^C \sum_{i \in \mathcal{I}_g} P_{g,i}^{max} x_{g,i} \right. \right.$$
$$\left. \left. - \sum_{i \in \mathcal{I}_g} x_{g,i} IC_{g,i} - \sum_{t \in \mathcal{T}} \Delta T_t MC_g q_{g,t} \right\} \right\}$$

Adding the constants $0 = \pi^C(\sum_{g\in\mathcal{G}}\sum_{i\in\mathcal{I}_g} P^{max}_{g,i}x^{**}_{g,i} - \sum_{g\in\mathcal{G}}\sum_{i\in\mathcal{I}_g} P^{max}_{g,i}x^{**}_{g,i})$ and $\sum_{g\in\mathcal{G}}(-\sum_{t\in\mathcal{T}}\Delta T_t\pi^M_t q^{**}_{g,t} + \sum_{i\in\mathcal{I}_g} x^{**}_{g,i}IC_{g,i} + \sum_{t\in\mathcal{T}}\Delta T_t MC_g q^{**}_{g,t})$ leads to equation (5.15). □

Proof of Proposition 5.7. We denote by $LOC_g(\pi^C, \pi^M)$ the lost opportunity cost of g under both energy and capacity prices. $LOC_g(0, \pi^M)$ then corresponds to the LOC in the energy-only market. Denoting the optimal objective function of equation (5.15) by ξ, since $\pi^C = 0$ is a feasible solution of the optimization problem (5.15), we conclude that: $\xi \leq 0 \times (\sum_g\sum_i P^{max}_{g,i}x^{**}_{g,i} - C^{min}) + \sum_g LOC^{gen}_g(0,\pi^M) = \sum_g LOC^{gen}_g(0,\pi^M)$. Furthermore, if $C^{min} \leq \sum_g\sum_i P^{max}_{g,i}x^*_{g,i}$ then $\pi^{C*}(\sum_g\sum_i P^{max}_{g,i}x^{**}_{g,i} - C^{min}) \geq 0$, from which we can write $\sum_g LOC^{gen}_g(\pi^{C*},\pi^M) \leq \xi$. We conclude that

$$\sum_g LOC^{gen}_g(\pi^{C*},\pi^M) \leq \sum_g LOC^{gen}_g(0,\pi^M)$$

□

5.B COMPREHENSIVE ERAA MATHEMATICAL MODEL

CONTINUOUS VANILLA ERAA MODEL. In this section we present the complete model of ERAA (see also Ávila et al. (2023)). The objective is to minimize the total cost:

$$\begin{aligned}\min \sum_{g\in\mathcal{G}^{new}} x^{new}_g IC^{new}_g - \sum_{g\in\mathcal{G}^{exist}} x^{exist}_g IC^{exist}_g + \Delta T \Bigg(\sum_{\substack{i\in\mathcal{N}\\ t\in\mathcal{T}}} \xi^+_{i,t} VOLL \\ + \sum_{\substack{t\in\mathcal{T}\\ g\in\mathcal{G}^{new}}} MC^{new}_g q^{new}_{g,t} + \sum_{\substack{t\in\mathcal{T}\\ g\in\mathcal{G}^{exist}}} MC^{exist}_g q^{exist}_{g,t} \\ + \sum_{\substack{t\in\mathcal{T}\\ g\in\mathcal{G}^{DSR}}} MC^{DSR}_{g,t} q^{DSR}_{g,t} + \sum_{\substack{t\in\mathcal{T},i\in\mathcal{N}\\ j\in from(i)}} |f_{i,j,t}| WC_{i,j} + \sum_{\substack{h\in\mathcal{H}\\ t\in\mathcal{T}}} s_{h,t} SP_h \Bigg)\end{aligned} \tag{5.21}$$

The market clearing condition is expressed as follows:

$$\begin{aligned}&\sum_{g\in\mathcal{G}^{new}_i} q^{new}_{g,t} + \sum_{g\in\mathcal{G}^{exist}_i} q^{exist}_{g,t} + \sum_{g\in\mathcal{G}^{DSR}_i} q^{DSR}_{g,t} + \sum_{g\in\mathcal{G}^{BAT}_i} (bd_{g,t} - bc_{g,t}) \\ &+ \xi^+_{i,t} - \xi^-_{i,t} + \sum_{h\in\mathcal{H}_i} q^{turb}_{h,t} - \sum_{h\in\mathcal{H}^{PS}_i} q^{pump}_{h,t} \\ &= D_{i,t} + \sum_{j\in from(i)} f_{i,j,t} - \sum_{j\in to(i)} f_{j,i,t} \quad \forall i\in\mathcal{N},\ \forall t\in\mathcal{T}\end{aligned} \tag{5.22}$$

The operational constraints of the assets are as follows:

$$0 \leq x_g^{new} \leq Capa_g^{max} \quad \forall g \in \mathcal{G}^{new} \tag{5.23a}$$

$$0 \leq x_g^{exist} \leq RCapa_g^{max} \quad \forall g \in \mathcal{G}^{exist} \tag{5.23b}$$

$$0 \leq q_{g,t}^{new} \leq x_g^{new} \quad \forall g \in \mathcal{G}^{new},\ t \in \mathcal{T} \tag{5.23c}$$

$$P_{g,t}^{min,exist} \leq q_{g,t}^{exist} \quad \forall g \in \mathcal{G}^{exist},\ t \in \mathcal{T} \tag{5.23d}$$

$$q_{g,t}^{exist} \leq P_{g,t}^{max,exist} - x_g^{exist} \frac{P_{g,t}^{max,exist}}{\max(P_{g,t}^{max,exist})} \quad \forall g \in \mathcal{G}^{exist},\ t \in \mathcal{T} \tag{5.23e}$$

$$0 \leq q_{g,t}^{DSR} \leq P_{g,t}^{max,DSR} \quad \forall g \in \mathcal{G}^{DSR},\ t \in \mathcal{T} \tag{5.23f}$$

$$F_{i,j,t}^{min} \leq f_{i,j,t} \leq F_{i,j,t}^{max} \quad \forall i \in \mathcal{N},\ j \in from(i),\ t \in \mathcal{T} \tag{5.23g}$$

$$0 \leq bd_{g,t}, bc_{g,t} \leq B_g^{max} \quad \forall g \in \mathcal{G}^{BAT}, t \in \mathcal{T} \tag{5.23h}$$

$$0 \leq bv_{g,t} \leq B_g^{capa} \quad \forall g \in \mathcal{G}^{BAT}, t \in \mathcal{T} \tag{5.23i}$$

$$bv_{g,t} = bv_{g,t-1} + \Delta T(B_g^{eff} bc_{g,t} - bd_{g,t}) \quad \forall g \in \mathcal{G}^{BAT},\ t \in \mathcal{T}^1 \tag{5.23j}$$

$$bv_{g,1} = B_g^{init} + \Delta T(B_g^{eff} bc_{g,1} - bd_{g,1}) \quad \forall g \in \mathcal{G}^{BAT} \tag{5.23k}$$

$$0 \leq v_{h,t}^{head} \leq V_h^{max} \quad \forall h \in \mathcal{H}^{res} \cup \mathcal{H}^{PSC} \cup \mathcal{H}^{PSO},\ t \in \mathcal{T} \tag{5.23l}$$

$$0 \leq q_{h,t}^{turb} \leq P_{h,t}^{max,turb} \quad \forall h \in \mathcal{H}^{res} \cup \mathcal{H}^{PSC} \cup \mathcal{H}^{PSO},\ t \in \mathcal{T} \tag{5.23m}$$

$$v_{h,t}^{head} = v_{h,t-1}^{head} + \Delta T(IF_{h,t} - q_{h,t}^{turb} - s_{h,t}) \quad \forall h \in \mathcal{H}^{res},\ t \in \mathcal{T}^1 \tag{5.23n}$$

$$v_{h,1}^{head} = V_h^{0,head} + \Delta T(IF_{h,1} - q_{h,1}^{turb} - s_{h,1}) \quad \forall h \in \mathcal{H}^{res} \tag{5.23o}$$

$$v_{h,t}^{head} = v_{h,t-1}^{head} + \Delta T(H_h^{eff} q_{h,t}^{pump} - q_{h,t}^{turb} - s_{h,t}) \quad \forall h \in \mathcal{H}^{PSC},\ t \in \mathcal{T}^1 \tag{5.23p}$$

$$v_{h,1}^{head} = V_h^{0,head} + \Delta T(H_h^{eff} q_{h,1}^{pump} - q_{h,1}^{turb} - s_{h,1}) \quad \forall h \in \mathcal{H}^{PSC} \tag{5.23q}$$

$$\begin{aligned} v_{h,t}^{head} &= v_{h,t-1}^{head} + \Delta T(IF_{h,t} + H_h^{eff} q_{h,t}^{pump} \\ &\quad - q_{h,t}^{turb} - s_{h,t}) \quad \forall h \in \mathcal{H}^{PSO},\ t \in \mathcal{T}^1 \end{aligned} \tag{5.23r}$$

$$v_{h,1}^{head} = V_h^{0,head} + \Delta T(IF_{h,1} + H_h^{eff} q_{h,1}^{pump} - q_{h,1}^{turb} - s_{h,1}) \quad \forall h \in \mathcal{H}^{PSO} \tag{5.23s}$$

$$v_{h,t}^{tail} = v_{h,t-1}^{tail} + \Delta T(-H_h^{eff} q_{h,t}^{pump} + q_{h,t}^{turb}) \quad \forall h \in \mathcal{H}^{PS},\ t \in \mathcal{T}^1 \tag{5.23t}$$

$$v_{h,1}^{tail} = V_h^{0,tail} + \Delta T(-H_h^{eff} q_{h,1}^{pump} + q_{h,1}^{turb}) \quad \forall h \in \mathcal{H}^{PS} \tag{5.23u}$$

$$0 \leq q_{h,t}^{pump} \leq P_{h,t}^{max,pump} \quad \forall h \in \mathcal{H}^{PS},\ t \in \mathcal{T} \tag{5.23v}$$

$$\xi_{i,t}^{+}, \xi_{i,t}^{-} \geq 0 \quad \forall i \in \mathcal{N},\ t \in \mathcal{T} \tag{5.23w}$$

The hydro technologies are of four types:

- *Run-of-River*: a turbine without any storage, i.e. essentially a natural inflow, which is directly accounted for in the net load.

- *Reservoir*: a turbine with a reservoir, fed with inflows, that enables to choose the turbine power at each hour.
- *PS Closed*: a first pumped-storage technology, composed of two reservoirs (head and tail) with pumps and turbines. The head reservoir is fed by pumped water. The tail reservoir is fed by turbine water. There are no natural inflows.
- *PS Open*: a second pumped-storage technology, which is the same as PS Closed, except that there are natural inflows feeding the head reservoir.

The model, although slightly more compact than the actual EVA model of ENTSO-E, includes *all* the features of the ENTSO-E model[27]. Table 5.7 provides the comprehensive nomenclature. We remark that D_t is the load net of RES production and run-of-river production. The parameter IC_g^{new} includes both annualized capital cost as well as fixed maintenance cost. The parameter IC_g^{exist} essentially includes fixed maintenance cost. The parameters MC_g^{exist} and MC_g^{new} include variable operation and maintenance cost, fuel cost as well as a CO_2 tax.

27 The main "simplifications" compared to the original EVA model are the following. (i) The HVDC and HVAC lines are merged together. This is justified by the fact that the network is represented as an ATC model. (ii) Certain parameters (such as *IC* or *MC*) are pre-processed in order to make the model more compact. (iii) Finally, the data in the Turkish zone leads to outlier results of investment in Turkey. This is also acknowledged by ENTSO-E (2021). Thus, we remove Turkey from the model. This is aligned with ENTSO-E's assumption in the 2022 study.

Sets	
$\mathcal{T}$, $\mathcal{T}^1$	set of periods and $\mathcal{T}^1 = \mathcal{T} \setminus \{1\}$
$\mathcal{N}$	set of nodes
$\mathcal{G}^{new}$, $\mathcal{G}_i^{new}$	set of new plants, set of new plants in node i
$\mathcal{G}^{exist}$, $\mathcal{G}_i^{exist}$	set of existing plants, set of existing plants in node i
$\mathcal{G}^{DSR}$, $\mathcal{G}_i^{DSR}$	set of DSR units, set of DSR units in node i
$\mathcal{G}^{BAT}$, $\mathcal{G}_i^{BAT}$	set of batteries, set of batteries in node i
$\mathcal{H}^{res}$, $\mathcal{H}^{PSC}$, $\mathcal{H}^{PSO}$	set of hydro units of type Reservoir, PS Closed & PS Open
$\mathcal{H}$, $\mathcal{H}^{PS}$	set of all the hydro and $\mathcal{H}^{PS} = \mathcal{H}^{PSC} \cup \mathcal{H}^{PSO}$
Parameters	
ΔT	duration of a time period
D_t	net load
$Capa_g^{max}$	max capacity that can be built
$RCapa_g^{max}$	max capacity that can be retired
IC_g^{new}	investment cost of a new plant
IC_g^{exist}	investment cost of an existing plant
MC_g^{new}	operating cost of a new plant
MC_g^{exist}	operating cost of an existing plant
$MC_{g,t}^{DSR}$	price for the demand response
$P_{g,t}^{min,exist}$	min production of an existing plant
$P_{g,t}^{max,exist}$	max production of an existing plant
$P_{g,t}^{max,DSR}$	max production of a DSR plant
$F_{i,j,t}^{max}$	max flow of line from i to j
$F_{i,j,t}^{min}$	min flow of line from i to j
$WC_{i,j}$	cost of flowing power from i to j
B_g^{eff}	charging/discharging efficiency of the battery
B_g^{capa}, B_g^{max}	battery volume capacity, charging/discharging capacity
B_g^{init}	battery initial volume (in $t = 0$)
SP_h	"spill penalty" for spilling water out of the reservoir
$IF_{h,t}$	natural inflows of water (Reservoir and PS Open)
V_h^{max}	max volume of the head reservoir
$V_h^{head,0}$	initial volume of the head reservoir
$V_h^{tail,0}$	initial volume of the tail reservoir
$P_{h,t}^{max,turb}$	max power for turbine

$P_{h,t}^{max,pump}$	max power for pump (PS Open/Closed)
H_h^{eff}	efficiency of pumping (PS Open/Closed)
Variables	
x_g^{new}	new capacity built
x_g^{exist}	capacity being retired
$q_{g,t}^{new}$	production of a new plant
$q_{g,t}^{exist}$	production of an existing plant
$q_{g,t}^{DSR}$	production of the demand response
$f_{i,j,t}$	flow of the line from i to j
$bv_{g,t}, bc_{g,t}, bd_{g,t}$	battery volume, charge and discharge
$q_{h,t}^{turb}, q_{h,t}^{pump}$	turbine and pump power
$v_{h,t}^{head}$	stored volume of water in head reservoir
$v_{h,t}^{tail}$	stored volume of water in tail reservoir
$s_{h,t}$	spilled volume
$\xi_{i,t}^{+}, \xi_{i,t}^{-}$	load and production shedding

TABLE 5.7: Nomenclature of problem (5.23).

DISCRETE INVESTMENT MODEL. The previous model is adapted as follows. Variables x_g^{new} and x_g^{exist} are non-negative integers: $x_g^{new}, x_g^{exist} \in \mathbb{N}$. The lumps of capacity—or power plant sizes—are modelled by parameters C_g^{new} and C_g^{exist}. The data for these parameters is provided in Table 5.8. The sole changes with respect to the comprehensive model (5.23) are the constraints on investment limits (equations (5.23a)-(5.23b)), the constraints on production limits (equations (5.23c)-(5.23e)) as well as the fixed cost term in the objective (5.21). The investment and production constraints are now expressed as follows:

$$0 \leq C_g^{new} x_g^{new} \leq Capa_g^{max} \quad \forall g \in \mathcal{G}^{new} \tag{5.24a}$$

$$0 \leq C_g^{exist} x_g^{exist} \leq RCapa_g^{max} \quad \forall g \in \mathcal{G}^{exist} \tag{5.24b}$$

$$0 \leq q_{g,t}^{new} \leq C_g^{new} x_g^{new} \quad \forall g \in \mathcal{G}^{new},\ t \in \mathcal{T} \tag{5.24c}$$

$$P_{g,t}^{min,exist} \leq q_{g,t}^{exist} \quad \forall g \in \mathcal{G}^{exist},\ t \in \mathcal{T} \tag{5.24d}$$

$$q_{g,t}^{exist} \leq P_{g,t}^{max,exist} - \frac{x_g^{exist} C_g^{exist} P_{g,t}^{max,exist}}{\max(P_{g,t}^{max,exist})} \quad \forall g \in \mathcal{G}^{exist},\ t \in \mathcal{T} \tag{5.24e}$$

Technology	Plant Size [MW]	Technology	Plant Size [MW]
Nuclear	1000	Light oil	100
Gas/CCGT new	500	Heavy oil/old 2	300
Gas/CCGT old 2	400	Heavy oil/old 1	200
Gas/CCGT present 2	450	Oil shale/new	250
Gas/CCGT present 1	450	Hard coal/new	600
Gas/CCGT old 1	400	Hard coal/old 1	550
Gas/OCGT new	300	Hard coal/old 2	800
Gas/OCGT old	250	Lignite/new	300
Gas/conventional old 1	200	Lignite/old 1	800
Gas/conventional old 2	200	Lignite/old 2	500

TABLE 5.8: Capacity lumps—or plant size—for the different technologies.

The fixed costs of the objective are:

$$\sum_{g\in\mathcal{G}^{new}} x_g^{new} C_g^{new} IC_g^{new} - \sum_{g\in\mathcal{G}^{exist}} x_g^{exist} C_g^{exist} IC_g^{exist} \tag{5.25}$$

The energy prices $\pi_{i,t}^M$ in this model are assumed to be the merit order prices of Definition 5.1. The investment decisions are fixed to their optimum $(x_g^{new,**}, x_g^{exist,**})$ and the prices are then obtained as the dual variables of the market clearing constraints.

CAPACITY MARKET. The capacity auction model reads as follows.

$$\min_{\substack{p\geq 0\\ x\in\mathbb{N}}} \sum_{g\in\mathcal{G}^{new}} \left(x_g^{new} C_g^{new} IC_g^{new} - \Delta T \sum_{t\in\mathcal{T}} (\pi_{i(g),t}^M - MC_g^{new}) q_{g,t}^{new} \right) \tag{5.26a}$$

$$+ \sum_{g\in\mathcal{G}^{exist}} \left(-x_g^{exist} C_g^{exist} IC_g^{exist} - \Delta T \sum_{t\in\mathcal{T}} (\pi_{i(g),t}^M - MC_g^{exist}) q_{g,t}^{exist} \right) \tag{5.26b}$$

$$\sum_{g\in\mathcal{G}_i^{new}} x_g^{new} C_g^{new} - \sum_{g\in\mathcal{G}_i^{exist}} x_g^{exist} C_g^{exist} \geq C_i^{min} \tag{5.26c}$$

$$0 \leq x_g^{new} \leq \lfloor Capa_g^{max} / C_g^{new} \rfloor \quad \forall g \in \mathcal{G}^{new} \tag{5.26d}$$

$$0 \leq x_g^{exist} \leq \lfloor RCapa_g^{max} / C_g^{exist} \rfloor \quad \forall g \in \mathcal{G}^{exist} \tag{5.26e}$$

$$0 \leq q_{g,t}^{new} \leq C_g^{new} x_g^{new} \quad \forall g \in \mathcal{G}^{new},\ t \in \mathcal{T} \tag{5.26f}$$

$$P_{g,t}^{min,exist} \leq q_{g,t}^{exist} \quad \forall g \in \mathcal{G}^{exist},\ t \in \mathcal{T} \tag{5.26g}$$

$$q_{g,t}^{exist} \leq P_{g,t}^{max,exist} - \frac{x_g^{exist} C_g^{exist} P_{g,t}^{max,exist}}{\max(P_{g,t}^{max,exist})} \quad \forall g \in \mathcal{G}^{exist},\ t \in \mathcal{T} \tag{5.26h}$$

The capacity prices π_i^C are the Lagrangian multipliers associated to constraint (5.26c).

5.C DETAILED NUMERICAL RESULTS

The detailed results of the summary Tables 5.3 and 5.4 are provided respectively in Tables 5.9 and 5.10. Table 5.9 also provides the correspondence between the scenario labels (e.g. 2025/7), used in the text of the article to denote the scenarios, and the climate years (e.g. 1989) used in the raw data of ENTSO-E. Let us notice that among the 35 scenarios (climate years) provided by ENTSO-E, we were not computationally able to solve 4 of them (the climate years 1988, 2000, 2005, 2006) which are, therefore, not reported in the tables.

Climate	Scenario	Total Cost			Commissioning		Decommissioning		LOC
Year	Label	Cont.	Disc.	Inc.	Cont.	Disc.	Cont.	Disc.	Disc.
1982	2025/1	7.472e10	7.493e10	0.3%	4811	5300	38233	35500	1.3032e9
1983	2025/2	7.427e10	7.453e10	0.4%	12029	11900	38020	34900	8.93e8
1984	2025/3	7.563e10	7.585e10	0.3%	4331	3500	33497	29150	1.7367e9
1985	2025/4	8.259e10	8.272e10	0.2%	23579	23900	10841	7950	1.4379e9
1986	2025/5	7.738e10	7.752e10	0.2%	10641	9600	18380	14650	8.005e8
1987	2025/6	8.214e10	8.23e10	0.2%	16320	16200	20120	16400	1.1895e9
1989	2025/7	7.385e10	7.409e10	0.3%	4745	3800	37790	33000	4.91e8
1990	2025/8	7.097e10	7.12e10	0.3%	1829	1300	35663	30750	2.5412e8
1991	2025/9	7.764e10	7.783e10	0.2%	4927	4300	23674	20550	1.207e9
1992	2025/10	7.406e10	7.427e10	0.3%	3445	3300	27520	23950	1.1151e9
1993	2025/11	7.624e10	7.644e10	0.3%	4316	3800	21012	17050	4.6549e8
1994	2025/12	7.353e10	7.38e10	0.4%	6640	5800	45503	40800	1.3519e9
1995	2025/13	7.346e10	7.372e10	0.4%	4967	4300	35929	32250	1.4473e9
1996	2025/14	7.982e10	8.001e10	0.2%	10150	9900	30945	27150	5.871e8
1997	2025/15	7.719e10	7.736e10	0.2%	6822	6800	33530	30650	3.831e9
1998	2025/16	7.472e10	7.491e10	0.3%	5600	5500	25829	23100	7.489e8
1999	2025/17	7.463e10	7.485e10	0.3%	8588	7300	32784	29000	3.2773e9
2001	2025/18	7.562e10	7.581e10	0.3%	4338	4300	29388	26700	5.246e8
2002	2025/19	7.471e10	7.491e10	0.3%	5841	5400	32895	29550	1.0566e9
2003	2025/20	8.017e10	8.04e10	0.3%	14713	13300	35522	29750	7.432e8

2004	2025/21	7.72e10	7.735e10	0.2%	7019	6300	25367	22700	5.307e8
2007	2025/22	7.214e10	7.238e10	0.3%	5689	6300	37386	34150	7.997e8
2008	2025/23	7.256e10	7.29e10	0.5%	3136	1800	46552	41300	1.4014e9
2009	2025/24	7.763e10	7.782e10	0.2%	6359	5900	18435	15750	1.383e9
2010	2025/25	8.358e10	8.376e10	0.2%	11713	10600	14991	10400	7.862e8
2011	2025/26	7.554e10	7.574e10	0.3%	6823	6800	21109	17950	2.743e9
2012	2025/27	7.811e10	7.823e10	0.2%	13708	13200	17311	14300	2.5346e8
2013	2025/28	7.75e10	7.769e10	0.2%	5919	4500	24448	20550	1.3712e9
2014	2025/29	7.228e10	7.258e10	0.4%	3690	3300	46929	43100	4.254e8
2015	2025/30	7.328e10	7.347e10	0.3%	5750	5000	33932	30350	6.574e8
2016	2025/31	7.708e10	7.725e10	0.2%	5740	5300	22735	20100	5.043e8
Average		7.614e10	7.634e10	0.3%	7554	7048	29560	25920	1.1392e9

TABLE 5.9: Comparison of the discrete and continuous results of ERAA.

Climate Year		Without capacity market			With capacity market		
		New units	Exist units	Total	Inelastic	Elastic	No Coord.
1982	*LOC*	3.396e8	9.636e8	1.303e9	1.651e8	2.603e8	9.988e8
	RS	1.926e8	9.6e8	1.153e9	1.468e7	2.808e7	4.155e7
	FO	1.469e8	3.521e6	1.505e8	1.505e8	2.322e8	9.573e8
1983	*LOC*	7.811e8	1.119e8	8.929e8	8.123e8	9.12e8	1.73e9
	RS	7.601e7	5.74e6	8.175e7	1.126e6	2.54e7	1.952e7
	FO	7.05e8	1.061e8	8.112e8	8.112e8	8.866e8	1.711e9
1984	*LOC*	6.647e8	1.072e9	1.737e9	6.055e8	8.153e8	1.87e9
	RS	7.247e7	1.068e9	1.14e9	8.536e6	3.135e7	3.422e7
	FO	5.923e8	4.668e6	5.969e8	5.969e8	7.839e8	1.835e9
1985	*LOC*	7.911e8	6.468e8	1.438e9	2.569e7	1.115e8	3.759e8
	RS	7.911e8	6.259e8	1.417e9	4.821e6	4.821e6	6.5e7
	FO	0.0	2.087e7	2.087e7	2.087e7	1.067e8	3.109e8
1986	*LOC*	4.804e8	3.201e8	8.005e8	5.978e8	7.812e8	1.089e9
	RS	5.245e7	1.588e8	2.113e8	0.0	0.0	1.398e6
	FO	4.28e8	1.612e8	5.892e8	5.978e8	7.812e8	1.088e9
1987	*LOC*	9.443e8	2.452e8	1.19e9	8.249e8	8.979e8	1.378e9
	RS	1.939e8	1.864e8	3.803e8	1.486e7	1.486e7	0.0
	FO	7.504e8	5.885e7	8.092e8	8.1e8	8.831e8	1.378e9
	LOC	3.534e8	1.376e8	4.91e8	4.863e8	6.354e8	1.144e9

1989	*RS*	1.802e7	0.0	1.802e7	1.335e7	4.154e7	3.244e7
	FO	3.354e8	1.376e8	4.73e8	4.73e8	5.939e8	1.111e9
	LOC	3.842e7	2.157e8	2.541e8	8.317e7	2.532e8	1.143e9
1990	*RS*	3.842e7	1.477e8	1.861e8	1.4e7	3.874e7	3.45e7
	FO	0.0	6.802e7	6.802e7	6.917e7	2.145e8	1.109e9
	LOC	8.765e8	3.305e8	1.207e9	1.122e9	1.301e9	1.756e9
1991	*RS*	2.387e7	8.05e7	1.044e8	1.975e7	2.662e7	2.662e7
	FO	8.526e8	2.5e8	1.103e9	1.103e9	1.274e9	1.73e9
	LOC	2.576e8	8.575e8	1.115e9	1.835e8	3.493e8	1.078e9
1992	*RS*	1.111e8	8.385e8	9.496e8	1.802e7	2.621e7	1.681e7
	FO	1.465e8	1.897e7	1.655e8	1.655e8	3.231e8	1.061e9
	LOC	3.953e8	7.019e7	4.655e8	3.965e8	5.583e8	1.386e9
1993	*RS*	3.947e7	4.276e7	8.223e7	1.108e7	1.044e7	1.044e7
	FO	3.558e8	2.743e7	3.833e8	3.854e8	5.479e8	1.376e9
	LOC	1.106e9	2.459e8	1.351e9	1.302e9	1.52e9	2.197e9
1994	*RS*	2.619e7	4.576e7	7.195e7	2.291e7	4.8e7	2.171e7
	FO	1.079e9	2.001e8	1.279e9	1.279e9	1.472e9	2.176e9
	LOC	6.289e8	8.184e8	1.447e9	6.073e8	7.575e8	1.371e9
1995	*RS*	6.91e7	7.831e8	8.522e8	1.219e7	4.209e7	3.381e7
	FO	5.598e8	3.527e7	5.951e8	5.951e8	7.154e8	1.337e9
	LOC	4.125e8	1.746e8	5.871e8	3.188e8	5.896e8	9.255e8
1996	*RS*	1.709e8	1.076e8	2.785e8	8.699e6	2.36e7	2.481e7
	FO	2.416e8	6.699e7	3.086e8	3.101e8	5.66e8	9.007e8
	LOC	2.769e9	1.062e9	3.83e9	3.774e9	3.779e9	4.016e9
1997	*RS*	2.477e7	6.559e7	9.036e7	3.421e7	3.421e7	2.122e7
	FO	2.744e9	9.961e8	3.74e9	3.74e9	3.745e9	3.995e9
	LOC	4.625e8	2.864e8	7.489e8	3.582e8	5.41e8	9.477e8
1998	*RS*	1.08e8	2.826e8	3.906e8	0.0	0.0	3.261e7
	FO	3.545e8	3.734e6	3.582e8	3.582e8	5.41e8	9.151e8
	LOC	2.545e9	7.323e8	3.278e9	3.267e9	3.466e9	3.697e9
1999	*RS*	1.915e7	1.555e7	3.47e7	2.379e7	2.573e7	1.554e7
	FO	2.526e9	7.168e8	3.243e9	3.243e9	3.44e9	3.681e9
	LOC	3.046e8	2.2e8	5.246e8	3.033e8	3.967e8	9.833e8
2001	*RS*	3.822e7	1.957e8	2.339e8	1.261e7	2.798e7	2.798e7
	FO	2.664e8	2.431e7	2.907e8	2.907e8	3.687e8	9.553e8
	LOC	2.839e8	7.727e8	1.057e9	4.977e8	6.292e8	1.423e9

2002	*RS*	2.321e7	5.572e8	5.804e8	2.137e7	2.276e7	2.2e7
	FO	2.607e8	2.155e8	4.762e8	4.763e8	6.065e8	1.401e9
	LOC	6.128e8	1.304e8	7.432e8	4.132e8	6.189e8	1.462e9
2003	*RS*	2.435e8	1.207e8	3.642e8	2.457e7	4.948e7	3.917e7
	FO	3.693e8	9.747e6	3.79e8	3.886e8	5.694e8	1.423e9
	LOC	3.155e8	2.152e8	5.307e8	5.307e8	6.39e8	1.086e9
2004	*RS*	1.017e7	0.0	1.017e7	1.017e7	1.017e7	1.017e7
	FO	3.053e8	2.152e8	5.205e8	5.205e8	6.288e8	1.076e9
	LOC	1.494e8	6.503e8	7.997e8	1.162e7	2.079e8	7.412e8
2007	*RS*	1.494e8	6.503e8	7.997e8	1.162e7	3.714e7	5.302e7
	FO	0.0	0.0	0.0	0.0	1.708e8	6.882e8
	LOC	1.013e9	3.884e8	1.401e9	1.319e9	1.544e9	2.431e9
2008	*RS*	0.0	1.316e8	1.316e8	4.958e7	5.062e7	1.824e7
	FO	1.013e9	2.568e8	1.27e9	1.27e9	1.494e9	2.413e9
	LOC	1.21e9	1.73e8	1.383e9	1.353e9	1.565e9	1.846e9
2009	*RS*	2.072e7	1.108e7	3.18e7	1.968e6	1.968e6	0.0
	FO	1.19e9	1.619e8	1.352e9	1.352e9	1.563e9	1.846e9
	LOC	1.196e8	6.666e8	7.862e8	1.029e8	1.238e9	6.17e8
2010	*RS*	3.189e7	6.62e8	6.939e8	1.012e7	0.0	1.012e7
	FO	8.771e7	4.592e6	9.23e7	9.274e7	1.238e9	6.068e8
	LOC	7.27e8	2.016e9	2.743e9	4.947e8	5.7e8	1.059e9
2011	*RS*	2.456e8	2.016e9	2.262e9	1.331e7	1.331e7	6.36e7
	FO	4.814e8	0.0	4.814e8	4.814e8	5.567e8	9.954e8
	LOC	1.545e8	9.896e7	2.535e8	5.663e7	1.929e8	4.525e8
2012	*RS*	1.139e8	9.277e7	2.067e8	9.862e6	9.862e6	0.0
	FO	4.058e7	6.192e6	4.677e7	4.677e7	1.83e8	4.525e8
	LOC	1.19e9	1.812e8	1.371e9	1.144e9	1.269e9	1.714e9
2013	*RS*	4.613e7	1.812e8	2.273e8	104000.0	1.576e7	1.69e7
	FO	1.144e9	0.0	1.144e9	1.144e9	1.253e9	1.697e9
	LOC	1.052e8	3.202e8	4.254e8	5.447e7	2.678e8	1.328e9
2014	*RS*	6.925e7	3.171e8	3.864e8	1.312e7	4.362e7	3.012e7
	FO	3.599e7	3.061e6	3.905e7	4.135e7	2.242e8	1.297e9
	LOC	3.969e8	2.605e8	6.574e8	5.8e8	7.717e8	1.232e9
2015	*RS*	9.872e6	1.15e8	1.248e8	4.744e7	1.837e7	1.931e7
	FO	3.871e8	1.455e8	5.326e8	5.326e8	7.533e8	1.213e9
	LOC	3.423e8	1.62e8	5.043e8	5.043e8	5.993e8	8.232e8

2016	*RS*	1.013e7	0.0	1.013e7	1.013e7	1.013e7	1.013e7
	FO	3.322e8	1.62e8	4.942e8	4.942e8	5.891e8	8.131e8
	LOC	6.7e8	4.692e8	1.139e9	7.192e8	9.044e8	1.429e9
Av.	*RS*	9.805e7	3.376e8	4.356e8	1.477e7	2.364e7	2.429e7
	FO	5.72e8	1.316e8	7.037e8	7.045e8	8.808e8	1.405e9

Table 5.10: Analysis of the agent incentives decomposed into lost opportunity costs (*LOC*), revenue shortfall (*RS*) and foregone opportunity (*FO*), for the two cases including or not a capacity payment. The results report three CRM settings: the inelastic capacity target, the elastic capacity demand curve and the inelastic capacity target computed without European coordination.

6

CONCLUSION

6.1 Summary of the Contributions

This thesis has studied what is often termed as "[. . .] one of the fundamental assumptions of microeconomics" (Mas-Colell et al., 1995, p. 133), the *convexity* of the market. Electricity markets provide a unique playing field for studying the impacts of relaxing this assumption. Indeed, since the implementation of the deregulation policies that led to the liberalization of the power sector, wholesale electricity markets have been characterized by the presence of non-convex bids. Therefore, the question of how to price these non-convexities, on top of being a fundamental and theoretical inquiry, has also taken a very practical form, attracting the interest of both academics and practitioners.

In chapter 3, we perform a theoretical and numerical analysis of different price formation rules proposed in the literature. We analyse six pricing methods, we establish several mathematical properties and we illustrate our findings on stylized examples and numerical simulations with realistic datasets. Although our analysis applies to both European and American electricity markets, it has been particularly motivated by the recent discussions among European stakeholders to reform the price formation rules adopted almost two decades ago by the "Trilateral Market Coupling" (2006). Our analysis highlights several advantages of convex hull pricing (CHP):

- CHP considerably reduces the lost opportunity costs, therefore improving the incentives faced by the market participants, mitigating the self-scheduling opportunities.

- This reduction is such that, not only the *total*, but the *distribution* of the lost opportunity costs across suppliers is significantly improved.
- CHP guarantees a theoretical bound on the total lost opportunity costs, which is independent from the number of market participants. This translates into a relative magnitude of lost opportunity costs with respect to the total surplus that shrinks with the market size (cf. Proposition 3.11).
- CHP ensures the consistency between cost minimization (or surplus maximization) and lost opportunity costs minimization, thus avoiding any clash between these two objectives.
- CHP also leads to modest revenue shortfalls. This is due to the fact that revenue shortfalls are a particular type of lost opportunity costs, where the "lost opportunity" is to self-schedule at 0. We further observe a *remarkable asymmetry*: while minimizing the lost opportunity costs (as CHP does) leads to low revenue shortfalls ; minimizing the revenue shortfalls (as MMWP does) can exacerbate the lost opportunity costs dramatically.
- These features are numerous advantages with respect to the alternative pricing rules. In particular, with respect to IP pricing, the fact that CHP incorporates the lumpy costs in the price signal improves significantly the incentives faced by the market agents. If ELMP would be a significant first step in the direction of CHP, it does not safeguard all the theoretical guarantees of CHP nor does it achieve the same performance in terms of lost opportunity costs minimization. Finally, while minimizing the revenue shortfall—or "make-whole payments"—may sound like a reasonable target, our analysis shows that approaches such as "MMWP pricing" may also result in unbearable lost opportunity costs.

Chapter 4 goes on with the computational challenges related to convex hull pricing. Indeed, if it has often been contemplated as a promising approach to price electricity in non-convex wholesale markets, a practical concern with CHP is that it turns out to be computationally challenging to calculate. In this chapter, we propose a dual decomposition algorithm known as the Level Method to efficiently compute locational convex hull prices. We adapt the basic algorithm to the specificities of convex hull pricing and we provide empirical evidence about the favorable performance of our algorithm on large test instances based on PJM and Central Europe markets (with respect to the number of suppliers, our largest dataset includes 1000 power units; with respect to the dimension of the price-space, our largest dataset includes 96 periods and 30 to 59 bidding zones, comparable to the size of the European day-ahead market). Our analysis shows that:

- The Level Method is able to compute convex hull prices on realistic-size datasets in a computing time compatible with the timing constraints of the wholesale markets. On a personal computer, the largest instance is solved in less than 10 minutes.
- On the so-called CWE "BE dataset", that includes 96 periods and 30 bidding zones, the Level Method is *5 times faster than the state-of-the art* in the literature.
- These results are hardly sensitive towards the choice of the "α" in the projection program, the main parameter of the Level Method.
- The Level Method exhibits a convergence path that is robust: it reaches near-optimal convex hull prices within a few iterations.

Chapter 5 analyses a source of non-convexities that is not discussed as broadly as the non-convexities in day-ahead power auctions: the indivisibilities in investment decisions. We study this problem by means of a capacity expansion model with indivisibilities. The main findings are illustrated with a numerical experiment conducted on the capacity expansion model used by ENTSO-E to assess the adequacy of the entire European system. Our analysis leads to several conclusions:

- Under indivisible investment decisions, a decentralized market for energy fails to reproduce the efficient investment plan. This failure can be measured by the concept of *long-term* lost opportunity costs.
- A popular argument to neglect indivisibilities is based on the smoothing effect induced by market size. Our analysis shows that, as far as the investment problem is concerned, under the classic "merit order pricing", this smoothing effect may not happen, consequently the long-term lost opportunity costs could be arbitrary large. The theoretical argument is confirmed by our numerical simulations.
- A capacity market that clears discrete offers could mitigate the lumpiness problem. We particularly introduce the novel concept of convex hull pricing for capacity auctions. Our analysis shows that, similarly to CHP in short-term auctions, it can mitigate *long-term* lost opportunity costs. This is confirmed by our numerical simulations. These simulations nevertheless stress the limits of a capacity market: its

effect can be inconclusive—and even counter-productive—when the CRM is ill-designed.

6.2 A WORD ABOUT THE CHANGES INDUCED BY THE ENERGY TRANSITION

This thesis has said very little about global warming or the energy transition. This might sound like an unforgivable oversight, knowing that the production of electricity stands for $\sim$25% of global anthropogenic greenhouse gas emissions (IPCC, 2014)[1]. In this concluding section, I would like to correct this omission by, first, briefly discussing the relationship between environmental policies and electricity markets; and second, to discuss the impact of the energy transition on the subject matter of this thesis—pricing under non-convexities.

GLOBAL WARMING AND THE ELECTRICITY MARKET. Two striking contemporary phenomena are, on the one hand, the increasing reliance on the market in our societies as an institution used to coordinate many aspects of our social and economical spheres and to allocate resources; and on the other hand, the alarming evidence of global warming and the massive environmental harm resulting from human economic activities (O'Neill, 2016). One could legitimately wonder what is the relationship between these two phenomena. From the viewpoint of economics, the relationship is straightforward: global warming is not the consequence of *too many* markets, but of *too few* markets. It results from externalities of our economic activities on the environment: because these externalities are not priced, they are over-produced. The absence of a market for carbon virtually implies a price for emitting carbon of 0€/tCO_2, which fails to signal the scarcity of this resource, resulting into inefficiencies.

If the environmental policies dictated by these economical principles are conceptually straight—Pigouvian taxation of carbon externalities—, implementing the right carbon tax, however, turns out to be both scientifically challenging and politically controversial. For example, the DICE model, developed by Nordhaus for more than thirty years, has significantly adapted its estimation of the social cost of carbon based on the scientific data collected along the years: it went from an estimate of the optimal carbon tax for 2015[2] of 5$/tCO2 with the DICE model of 1992, to an esti-

1 This figure is taken from the Fifth Assessment Report of IPCC and accounts for all anthropogenic greenhouse gas emissions, including the so-called "AFOLU". Let us also stress that this is a *global* figure. In a country like Belgium, that has already phased out coal-fired power plants, and that uses nuclear plants as base-load units, this number is likely much lower.

2 In money adjusted for inflation, in 2010$.

mate of 31\$/tCO2 with the DICE model of 2016.[3] At the same time, other approaches, such as the Stern Review, advocate for a lower discount rate and thus for a much higher tax of 100\$/tCO2, or even more[4].

Actual implementations of carbon price policies have nonetheless followed, although with a varying level of ambition. In 2023, 23% of emissions worldwide are covered by some sort of carbon price, the level of which varies significantly across countries[5]. In the US, the average national carbon price is 25\$/tCO_2 but it covers less than 10% of national emissions. In Europe, the Emissions Trading System (ETS), a cap-and-trade system, covers 40% of emissions and is close to ~90\$/tCO_2[6]. Another broadly adopted approach is the *subsidy-based* approach, instead (or on top) of a *taxed-based* (carbon price) approach. A recent example is the Inflation Reduction Act voted in 2022 in the US.

All-in-all, the consequences of these various policies for the electricity sector have been unprecedented. Two important effects are the phase-out of coal-fired units, which are carbon intensive, and the massive investments into renewable electricity plants such as wind and PV. Figure 6.1 illustrates these trends with the case of Belgium. The last units of coal were phased out a decade ago. As far as renewable energy resources are concerned, in 2012, wind and PV generation capacities stood for less than 20% of the Belgian power fleet. Ten years later, in 2023, this number has raised to almost 50%. Looking further in the past, back in 2006, there were 212MW and 2MW of, respectively, wind and PV. In 2023, these figures have raised to 5504MW and 8549MW[7].

ENERGY TRANSITION AND PRICING UNDER NON-CONVEXITIES. The shift from the "old" system, dominated by thermal generation units, to a system with a large share of renewable resources does not invalidate the basic market design principles that have been outlined in chapter 2. On the contrary, it makes them even more important and it makes the flaws in the implementation of these principles more critical (Hogan, 2019,

3 Nordhaus (2018) provides a comprehensive analysis of the changes between the first version of DICE in 1992 (Nordhaus, 1992) and the version of 2016: the social cost of carbon turns out to be one of the most severely flawed estimate of the 1992' model.

4 The carbon tax estimated with DICE, using the discount rate advocated by Stern is around 100\$/tCO2 for 2015 (Nordhaus, 2008).

5 For the following numbers, see Clausing and Wolfram (2023) who rely on data from the World Bank.

6 It is worth stressing the volatility of the ETS price, that reached ~100€/tCO_2 in early 2023, but dropped to ~50€/tCO_2 in early 2024.

7 These number are sourced from the "renewable energy capacity statistics" of IRENA

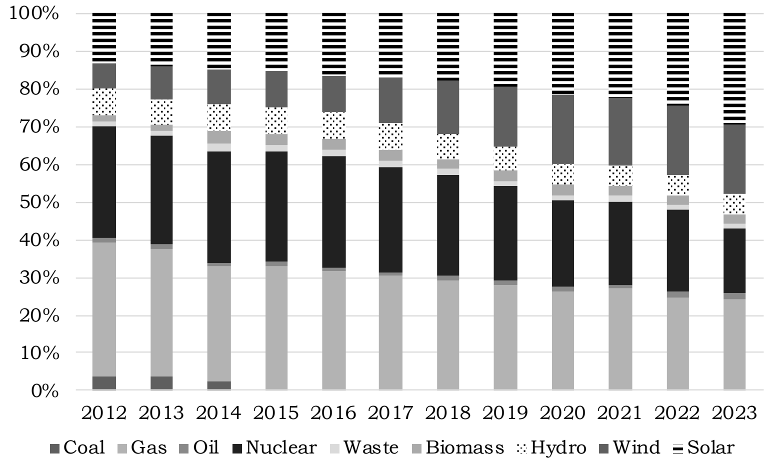


FIGURE 6.1: Electricity generation capacity in Belgium. [Data source: IRENA for wind and PV, Elia for the remaining technologies]

2022).[8] The massive integration of renewable energy sources, with typically high capital cost and low—if not zero—variable production cost, tends to lower the electricity price—as renewable resources are "zero-marginal cost" resources—, therefore reducing the surplus earned by the conventional power plants from the electricity prices. In this context, there is a pressing need to rightly price the *flexibility* offered by these "conventional power plants" that are making less profit from selling power and more by providing the system with their flexibility (Hogan, 2019).

As a consequence, one could argue that the topic of "pricing non-convexities" becomes even more important in the "new" system. Indeed, the penetration of renewable ressources renders the operations of thermal plants more cyclic, increasing the share of start-up costs in the variable production cost. As an example, Schill et al. (2017) estimates that the share of start-up costs in variable production cost of a CCGT in Germany unit has doubled between 2013 and 2020. Since the thermal units will have to switch on and off more frequently to compensate for the intermittence of renewable sources, accounting for their fixed costs and operational constraints in the market, and reflecting this cost of being *flexible*—start-up cost, but also ramping capability—in the price signal remains evermore critical.

8 "The impact of greater penetration of intermittent generation underscores the importance of continuing to improve implementation of the basic efficient market design without requiring a fundamental change in the underlying theory." (Hogan, 2022)

BIBLIOGRAPHY

Álvarez, C., Mancilla-David, F., Escalona, P., and Angulo, A. (2019). A Bienstock–Zuckerberg-based algorithm for solving a network-flow formulation of the convex hull pricing problem. *IEEE Transactions on Power Systems*, 35(3):2108–2119.

Andrianesis, P., Bertsimas, D., Caramanis, M., and Hogan, W. (2021). Computation of convex hull prices in electricity markets with non-convexities using dantzig-wolfe decomposition. *IEEE Transactions on Power Systems*, 37(4):2578–2589.

Aravena, I. and Papavasiliou, A. (2016). Renewable energy integration in zonal markets. *IEEE Transactions on Power Systems*, 32(2):1334–1349.

Arrow, K. J. (1962). The economic implications of learning by doing. *The review of economic studies*, 29(3):155–173.

Arrow, K. J. and Hahn, F. (1971). *General competitive analysis*. Holden-Day.

Ávila, D., Papavasiliou, A., Junca, M., and Exizidis, L. (2023). Applying high-performance computing to the european resource adequacy assessment. *IEEE Transactions on Power Systems*.

Belpex, APX, and Powernext (2006). Trilateral market coupling, algorithm appendix. https://inis.iaea.org/collection/NCLCollectionStore/_Public/38/045/38045712.pdf.

Bergheimer, S., Cantillon, E., and Reguant, M. (2023). Price and quantity discovery without commitment. *International Journal of Industrial Organization*, 90:102987.

Bertsekas, D., Lauer, G., Sandell, N., and Posbergh, T. (1983). Optimal short-term scheduling of large-scale power systems. *IEEE Transactions on Automatic Control*, 28(1):1–11.

Bertsekas, D. and Sandell, N. (1982). Estimates of the duality gap for large-scale separable nonconvex optimization problems. In *1982 21st IEEE conference on decision and control*, pages 782–785. IEEE.

Bichler, M., Knörr, J., and Maldonado, F. (2022). Pricing in nonconvex markets: How to price electricity in the presence of demand response. *Information Systems Research*, 34(2):652–675.

Birge, J. and Louveaux, F. (1988). A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34(3):384–392.

Birge, J. and Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Bohn, R., Caramanis, M., and Schweppe, F. (1984). Optimal pricing in electrical networks over space and time. *The Rand Journal of Economics*, pages 360–376.

Boiteux, M. (1960). Peak-load pricing. *The Journal of Business*, 33(2):157–179.

Borenstein, S. and Bushnell, J. (2000). Electricity restructuring: deregulation or reregulation. *Regulation*, 23:46.

Borenstein, S. and Bushnell, J. (2015). The US electricity industry after 20 years of restructuring. *Annual Review of Economics*, 7(1):437–463.

Borenstein, S., Bushnell, J., and Mansur, E. (2023). The economics of electricity reliability. *The Journal of Economic Perspectives*, 37(4):181–206.

Boyd, S. and Vandenberghe, L. (1997). Semidefinite programming relaxations of non-convex problems in control and combinatorial optimization. In *Communications, Computation, Control, and Signal Processing*, pages 279–287. Springer.

Boyd, S., Vandenberghe, L., and Skaf, J. (2008). Analytic center cutting-plane method. Stanford University, California, USA, Unpublished lecture notes, https://stanford.edu/class/ee364b/lectures/accpm_notes.pdf.

Byers, C. and Hug, G. (2022). Economic impacts of near-optimal solutions with non-convex pricing. *Electric Power Systems Research*, 211:108287.

Byers, C. and Hug, G. (2023). Long-run optimal pricing in electricity markets with non-convex costs. *European Journal of Operational Research*, 307(1):351–363.

CAISO (2020). Regulated Tariffs, Appendix C. http://www.caiso.com/Documents/AppendixC-LocationalMarginalPrice-asof-Sep9-2020.pdf.

Chao, H.-p. (2019). Incentives for efficient pricing mechanism in markets with non-convexities. *Journal of Regulatory Economics*, 56(1):33–58.

Chen, Y., O'Neill, R., and Whitman, P. (2020). A unified approach to solve convex hull pricing and average incremental cost pricing with large system study. *Working Paper*.

Clausing, K. A. and Wolfram, C. (2023). Carbon border adjustments, climate clubs, and subsidy races when climate policies vary. *Journal of Economic Perspectives*, 37(3):137–162.

Coase, R. H. (1937). The nature of the firm. *Economica*, 4(16):386–405.

Coase, R. H. (1946). The marginal cost controversy. *Economica*, 13(51):169–182.

Coase, R. H. (1970). The theory of public utility pricing and its application. *The Bell Journal of Economics and Management Science*, pages 113–128.

Coase, R. H. (1988). *The firm, the market, and the law*. University of Chicago press.

Commission Regulation (EU) (2015). Establishing a guideline on capacity allocation and congestion management (2015/1222). http://data.europa.eu/eli/reg/2015/1222/oj.

Cramton, P. (2017). Electricity market design. *Oxford Review of Economic Policy*, 33(4):589–612.

Cramton, P., Ockenfels, A., and Stoft, S. (2013). Capacity market fundamentals. *Economics of Energy & Environmental Policy*, 2(2):27–46.

Cramton, P. and Stoft, S. (2005). A capacity market that makes sense. *The Electricity Journal*, 18(7):43–54.

Damcı-Kurt, P., Küçükyavuz, S., Rajan, D., and Atamtürk, A. (2013). A polyhedral study of ramping in unit commitment. *Univ. California-Berkeley, Res. Rep. BCOL*, 13.

Davis, L. W. and Wolfram, C. (2012). Deregulation, consolidation, and efficiency: Evidence from us nuclear power. *American Economic Journal: Applied Economics*, 4(4):194–225.

De Maere d'Aertrycke, G., Ehrenmann, A., and Smeers, Y. (2017). Investment with incomplete markets for risk: The need for long-term contracts. *Energy Policy*, 105:571–583.

Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*, volume 17. Yale University Press.

Eldridge, B., O'Neill, R., and Hobbs, B. (2019). Near-optimal scheduling in day-ahead markets: pricing models and payment redistribution bounds. *IEEE transactions on power systems*, 35(3):1684–1694.

Elia (2019). Overview of belgian crm design: introduction note. https://www.elia.be/-/media/project/elia/elia-site/public-consultations/20190913/20190913_overview_crm_design.pdf.

Elia (2022). Capacity remuneration mechanism (crm). functioning rules. https://www.elia.be/-/media/project/elia/elia-site/users-group/ug/wg-adequacy/2022/20220513_crm_functioning_rules_clean_en.pdf.

ENTSO-E (2021). European resource adequacy assessment. Technical report, https://eepublicdownloads.azureedge.net/clean-documents/sdc-documents/ERAA/ERAA_2021_Executive%20Report.pdf.

EPRI (2019). Independent System Operator and Regional Transmission Organization Price Formation Working Group White Paper. Current Practice and Research Gaps in Alternative (Fast-Start) Price Formation Modeling. https://www.epri.com/research/products/3002013724.

Fabra, N. (2018). A primer on capacity mechanisms. *Energy Economics*, 75:323–335.

Fabrizio, K. R., Rose, N. L., and Wolfram, C. D. (2007). Do markets reduce costs? assessing the impact of regulatory restructuring on us electric generation efficiency. *American Economic Review*, 97(4):1250–1277.

Federal Energy Regulatory Commission (Apr. 2019). Docket No. EL18-34-000.

FERC (2014). Price Formation in Organized Wholesale Electricity Markets, Docket No. AD14-14-000. https://www.ferc.gov/sites/default/files/2020-05/AD14-14-operator-actions.pdf.

Finon, D. and Beeker, E. (2022). Le modèle d'acheteur central, une réponse aux défauts du marché électrique actuel. *La Revue de l'Énergie*, 662:47–61.

Frangioni, A. (2020). Standard bundle methods: untrusted models and duality. In *Numerical Nonsmooth Optimization*, pages 61–116. Springer.

Garcia, M., Nagarajan, H., and Baldick, R. (2020). Generalized convex hull pricing for the ac optimal power flow problem. *IEEE Transactions on Control of Network Systems*, 7(3):1500–1510.

Gentile, C., Morales-Espana, G., and Ramos, A. (2017). A tight mip formulation of the unit commitment problem with start-up and shut-down constraints. *EURO Journal on Computational Optimization*, 5(1):177–201.

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.

Gribik, P. R., Hogan, W. W., Pope, S. L., et al. (2007). Market-clearing electricity prices and energy uplift. *Cambridge, MA*.

Grubb, M. and Newbery, D. (2018). Uk electricity market reform and the energy transition: Emerging lessons. *The Energy Journal*, 39(6).

Harbord, D. and Pagnozzi, M. (2014). Britain's electricity capacity auctions: lessons from colombia and new england. *The Electricity Journal*, 27(5):54–62.

Helm, D. (2017). Cost of energy review. `www.dieterhelm.co.uk/energy/energy/cost-of-energyreview-independent-report/`.

Heuberger, C. F., Rubin, E. S., Staffell, I., Shah, N., and Mac Dowell, N. (2017). Power capacity expansion planning considering endogenous technology cost learning. *Applied Energy*, 204:831–845.

Hogan, W. W. (2002). Electricity market restructuring: reforms of reforms. *Journal of Regulatory Economics*, 21(1):103–132.

Hogan, W. W. (2016). Virtual bidding and electricity market design. *The Electricity Journal*, 29(5):33–47.

Hogan, W. W. (2019). Market design practices: Which ones are best?[in my view]. *IEEE Power and Energy Magazine*, 17(1):100–104.

Hogan, W. W. (2021). Strengths and weaknesses of the pjm market model. In Glachant, J.-M., Joskow, P., and Pollitt, M., editors, *Handbook on Electricity Markets*, pages 182–204. Edward Elgar Publishing.

Hogan, W. W. (2022). Electricity market design and zero-marginal cost generation. *Current Sustainable/Renewable Energy Reports*, pages 1–12.

Hogan, W. W. and Ring, B. J. (2003). On minimum-uplift pricing for electricity markets. *Electricity Policy Group*, pages 1–30.

Hua, B. and Baldick, R. (2017). A convex primal formulation for convex hull pricing. *IEEE Transactions on Power Systems*, 32(5):3814–3823.

Hutcheon, N. and Bialek, J. W. (2013). Updated and validated power flow model of the main continental european transmission network. In *2013 IEEE Grenoble Conference*, pages 1–5. IEEE.

IESO (2023). A progress report on contracted electricity supply. Technical report.

IEX (2020). Power Markets and Exchange Operations. https://eal.iitk.ac.in/assets/docs/Prasanna_Rao_Electricity%20Markets-IIT%20KANPUR.pdf.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.), Geneva, Switzerland.

Joskow, P. L. (2003). Electricity sector restructuring and competition: Lessons learned. *Cuadernos de economía*, 40(121):548–558.

Joskow, P. L. (2007). Competitive electricity markets and investment in new generating capacity. In Helm, D., editor, *The New Energy Paradigm*, chapter 4. Oxford University Press, Oxford.

Joskow, P. L. (2008). Lessons learned from electricity market liberalization. *The Energy Journal*, 29(Special Issue# 2).

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.

Joskow, P. L. (2022). From hierarchies to markets and partially back again in electricity: responding to decarbonization and security of supply goals. *Journal of Institutional Economics*, 18(2):313–329.

JPEX (2023). JPEX market data. https://www.jepx.jp/en/electricpower/market-data/spot/ave_year.html [Accessed: 2023-10-17].

Kahn, A. E., Cramton, P. C., Porter, R. H., and Tabors, R. D. (2001). Uniform pricing or pay-as-bid pricing: a dilemma for california and beyond. *The electricity journal*, 14(6):70–79.

Knueven, B., Ostrowski, J., and Watson, J.-P. (2020). On mixed-integer programming formulations for the unit commitment problem. *INFORMS Journal on Computing*, 32(4):857–876.

Krall, E., Higgins, M., and O'Neill, R. (2012). Rto unit commitment test system. *Federal Energy Regulatory Commission*, 98.

Lemaréchal, C. (2001). Lagrangian relaxation. In *Computational combinatorial optimization*, pages 112–156. Springer.

Lemaréchal, C. and Oustry, F. (2001). Sdp relaxations in combinatorial optimization from a lagrangian viewpoint. In *Advances in Convex Analysis and Global Optimization*, pages 119–134. Springer.

Lemaréchal, C. and Renaud, A. (2001). A geometric study of duality gaps, with applications. *Mathematical Programming*, 90(3):399–427.

Liberopoulos, G. and Andrianesis, P. (2016). Critical review of pricing schemes in markets with non-convex costs. *Operations Research*, 64(1):17–31.

Madani, M. and Papavasiliou, A. (2022). A note on a revenue adequate pricing scheme that minimizes make-whole payments. *18th International Conference on the European Energy Market (EEM)*, pages 1–6.

Madani, M., Ruiz, C., Siddiqui, S., and Van Vyve, M. (2018). Convex hull, ip and european electricity pricing in a european power exchanges setting with efficient computation of convex hull prices. *arXiv preprint arXiv:1804.00048*.

Marsten, R. E., Hogan, W. W., and Blankenship, J. W. (1975). The boxstep method for large-scale optimization. *Operations Research*, 23(3):389–405.

Mas-Colell, A., Whinston, M., and Green, J. (1995). *Microeconomic theory*. Oxford university press New York.

Mastropietro, P., Rodilla, P., and Batlle, C. (2017). Performance incentives in capacity mechanisms: Conceptual considerations and empirical evidence. *Economics of Energy & Environmental Policy*, 6(1):149–164.

Mays, J., Morton, D., and O'Neill, R. (2021). Investment effects of pricing schemes for non-convex markets. *European Journal of Operational Research*, 289(2):712–726.

MCSC (2022). Market Coupling Steering Committee: Market Coupling Consultative Group meeting, 1st of December. `https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/events/2022/MCCG-presentation-01122022final.pdf`.

MCSC (2023). 4th Market Coupling Consultative Group workshop, 20th of October. https://www.nemo-committee.eu/assets/files/MCCG-20102023-for-publishing-1bba37e3d4e7ca9ce1680954e91acc95.pdf.

Meeus, L. (2020). *The evolution of electricity markets in Europe*. Edward Elgar Publishing.

Mezghani, I., Stevens, N., Papavasiliou, A., and Chatzigiannis, D. I. (2023). Hierarchical coordination of transmission and distribution system operations in european balancing markets. *IEEE Transactions on Power Systems*, 38(5):3990–4002.

Morales-España, G., Gentile, C., and Ramos, A. (2015). Tight mip formulations of the power-based unit commitment problem. *OR Spectrum*, 37(4):929–950.

Morales-España, G., Latorre, J. M., and Ramos, A. (2013). Tight and compact milp formulation for the thermal unit commitment problem. *Power Systems, IEEE Transactions on*, 28(4):4897–4908.

N-SIDE (2021). Indian electricity prices determined with innovative belgian technology. https://energy.n-side.com/blog/indian-electricity-prices-determined-with-innovative-belgian-technology.

NEMO Committee (2019). Release note euphemia 10.3.

NEMO Committee (2020a). CACM annual report 2019.

NEMO Committee (2020b). Euphemia public description, single price coupling algorithm. Technical report, October.

NEMO Committee (2022). CACM annual report 2021. https://www.nemo-committee.eu/assets/files/nemo_CACM_Annual_Report_2021_220630-4e7321983974b812f28730a301c9f7d9.pdf.

NEMO Committee (2023). CACM annual report 2022. https://www.nemo-committee.eu/assets/files/cacm-annual-report-2022.pdf.

Nesterov, Y. (2004). *Introductory lectures on convex optimization: a basic course*. Springer.

Newbery, D. (2021). Designing an incentive-compatible efficient renewable electricity support scheme. *Energy Policy Research Group 2107, Cambridge Working Paper in Economics*.

Nordhaus, W. (1992). An optimal transition path for controlling greenhouse gases. *Science*, 258(5086):1315–1319.

Nordhaus, W. (2008). *A question of balance: Weighing the options on global warming policies*. Yale University Press.

Nordhaus, W. (2018). Evolution of modeling of the economics of global warming: changes in the dice model, 1992–2017. *Climatic change*, 148(4):623–640.

NYISO (2016). NYISO Hybrid Gas Turbine Pricing, June 27. https://www.ferc.gov/sites/default/files/2020-08/NYISO_Zhang.pdf.

O'Neill, J. (2016). Markets, ethics, and environment. *The Oxford Handbook of Environmental Ethics*, page 40.

O'Neill, R., Chen, Y., and Whitman, P. (2023). One-pass average incremental cost pricing. *Optimization Online*.

O'Neill, R. P., Sotkiewicz, P. M., Hobbs, B. F., Rothkopf, M. H., and Stewart Jr, W. R. (2005). Efficient market-clearing prices in markets with nonconvexities. *European journal of operational research*, 164(1):269–285.

Oren, S. S. (2000). Capacity payments and supply adequacy in competitive electricity markets. *Sepope, May*, 8.

Papavasiliou, A. (2021). Overview of EU capacity remuneration mechanisms. report for the Greek Regulatory Authority for Energy (RAE). Technical report.

Papavasiliou, A. (2024). *Optimization Models in Electricity Markets*. Cambridge University Press.

PJM (2017). Proposed enhancements to energy price formation.

Queyranne, M. and Wolsey, L. A. (2017). Tight mip formulations for bounded up/down times and interval-dependent start-ups. *Mathematical Programming*, 164(1):129–155.

Rajan, D. and Takriti, S. (2005). Minimum up/down polytopes of the unit commitment problem with start-up costs. *IBM Res. Rep.*

Roques, F. and Finon, D. (2017). Adapting electricity markets to decarbonisation and security of supply objectives: Toward a hybrid regime? *Energy Policy*, 105:584–596.

Scarf, H. (1994). The allocation of resources in the presence of indivisibilities. *Journal of Economic Perspectives*, 8(4):111–128.

Schill, W.-P., Pahle, M., and Gambardella, C. (2017). Start-up costs of thermal power plants in markets with increasing shares of variable renewable generation. *Nature Energy*, 2(6):1–6.

Schiro, D. A., Zheng, T., Zhao, F., and Litvinov, E. (2015). Convex hull pricing in electricity markets: Formulation, analysis, and implementation challenges. https://web.archive.org/web/20200320022732id_/http://www.optimization-online.org/DB_FILE/2015/03/4830.pdf.

Schmalensee, R. (2021). Strengths and weaknesses of traditional arrangements for electricity supply. In Glachant, J.-M., Joskow, P., and Pollitt, M., editors, *Handbook on Electricity Markets*, pages 13–35. Edward Elgar Publishing.

SDAC (2023). Non-uniform pricing: Explanatory note. https://www.nemo-committee.eu/assets/files/sdac-non-uniform-pricing-explanatory-note.pdf.

Silbernagl, M., Huber, M., and Brandenberg, R. (2015). Improving accuracy and efficiency of start-up cost formulations in mip unit commitment by modeling power plant temperatures. *IEEE Transactions on Power Systems*, 31(4):2578–2586.

Sioshansi, R., O'Neill, R., and Oren, S. S. (2008). Economic consequences of alternative solution methods for centralized unit commitment in day-ahead electricity markets. *IEEE Transactions on Power Systems*, 23(2):344–352.

Spees, K., Newell, S. A., and Pfeifenberger, J. P. (2013). Capacity markets—lessons learned from the first decade. *Economics of Energy & Environmental Policy*, 2(2):1–26.

Sridhar, S., Linderoth, J., and Luedtke, J. (2013). Locally ideal formulations for piecewise linear functions with indicator variables. *Operations Research Letters*, 41(6):627–632.

Starr, R. M. (1969). Quasi-equilibria in markets with non-convex preferences. *Econometrica*, pages 25–38.

Stevens, N. (2016). Models and algorithms for pricing electricity in unit commitment.

Stevens, N. and Papavasiliou, A. (2022). Application of the level method for computing locational convex hull prices. *IEEE Transactions on Power Systems*, 37(5):3958–3968.

Stevens, N., Papavasiliou, A., and Smeers, Y. (2024a). On some advantages of convex hull pricing for the european electricity auction. *Energy Economics*, 134:107542.

Stevens, N., Smeers, Y., and Papavasiliou, A. (2024b). Indivisibilities in investment and the role of a capacity market. *Journal of Regulatory Economics*, 66:238–272.

Stoft, S. (2002). *Power system economics: designing markets for electricity*, volume 468. IEEE press Piscataway.

Taylor, J. A. (2015). *Convex optimization of power systems*. Cambridge University Press.

UK Department of Energy & Climate Change (2013). Electricity Market Reform - Contract for Difference: Contract and Allocation Overview . https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/404463/EMR__Contract_for_Difference__Contract_and_Allocation_Overview_Final_28_August.pdf.

Van Vyve, M. (2011). Linear prices for non-convex electricity markets: models and algorithms. *CORE Discussion Paper 50*.

Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming. *SIAM review*, 38(1):49–95.

Vanderbeck, F. and Wolsey, L. A. (2010). Reformulation and decomposition of integer programs. In *50 Years of Integer Programming 1958-2008*, pages 431–502. Springer.

Wang, C., Luh, P. B., Gribik, P., Zhang, L., and Peng, T. (2009). A subgradient-based cutting plane method to calculate convex hull market prices. *Power & Energy Society General Meeting, 2009. PES'09. IEEE*, pages 1–7.

Wang, G., Shanbhag, U. V., Zheng, T., Litvinov, E., and Meyn, S. (2013a). An extreme-point subdifferential method for convex hull pricing in energy and reserve markets—Part I: Algorithm structure. *Power Systems, IEEE Transactions on*, 28(3):2111–2120.

Wang, G., Shanbhag, U. V., Zheng, T., Litvinov, E., and Meyn, S. (2013b). An extreme-point subdifferential method for convex hull pricing in energy and reserve markets—Part II: Convergence analysis and numerical performance. *Power Systems, IEEE Transactions on*, 28(3):2121–2127.

Williamson, O. E. (1966). Peak-load pricing and optimal capacity under indivisibility constraints. *The American Economic Review*, 56(4):810–827.

Wilson, R. (2002). Architecture of power markets. *Econometrica*, 70(4):1299–1340.

Wolsey, L. A. (1998). *Integer programming*. John Wiley & Sons.

Yu, Y., Guan, Y., and Chen, Y. (2020). An extended integral unit commitment formulation and an iterative algorithm for convex hull pricing. *IEEE Transactions on Power Systems*, 35(6):4335–4346.

Zhao, F., Schiro, D., Zhao, J., Zheng, T., and Litvinov, E. (2021). On the formulation dependence of convex hull pricing. *Optimization Online*.