Utilities Policy 81 (2023) 101488

Contents lists available at ScienceDirect

Utilities Policy

journal homepage: www.elsevier.com/locate/jup

Full-length article

Implementation of scarcity pricing without co-optimization in European energy-only balancing markets

Anthony Papavasiliou[a,*], Jacques Cartuyvels[b], Gilles Bertrand[c,1], Alain Marien[c,1]

[a] Department of Electrical and Computer Engineering at the National Technical University of Athens, Greece
[b] Center for Operations Research and Econometrics at UCLouvain, Belgium
[c] Belgian Regulatory Commission for Electricity and Gas (CREG), Belgium

ARTICLE INFO

Handling Editor: Janice A. Beecher

Keywords:
Scarcity pricing
Operating reserve demand curve
Balancing market

ABSTRACT

The implementation of scarcity pricing is underway in the Belgian balancing market. The market design proposed in this paper aims at transposing the first principles of scarcity pricing theory to the boundary conditions of European balancing markets. One relevant boundary condition is the absence of real-time co-optimization of energy and reserves in Europe. As international experience demonstrates, the fact that energy and reserves are not co-optimized in balancing markets does not preclude the implementation of scarcity pricing. The mechanism can be implemented implicitly, and a concrete path has been proposed in the context of the Belgian balancing market. The Belgian design proposal, which is based on the implicit trading of reserve in real time, has raised questions related to financial implications for market stakeholders, the pricing of multiple reserve products, cross-border interactions, the financing of the mechanism, compatibility with EU law, and the coexistence of the proposed mechanism with capacity markets. We attempt to address these questions in the present work by drawing comparisons between the explicit co-optimization of energy and reserves and the implicit trading of reserve capacity.

1. Context

Scarcity pricing refers to the practice or set of mechanisms that are in place in electricity markets for setting prices above the short-run variable costs of generation resources during periods of system stress.[2] In markets with price-responsive demand, shortage in production capacity results in price increases, which induce a demand-side response. This enables market forces to restore an equilibrium between supply and demand, and results in prices that allow producers to recover their long-run investment costs. Standard arguments (Boiteux, 1960) establish that this process results in a long-run equilibrium in the market that matches the outcome of a socially optimal expansion plan.

The absence of price elasticity in electricity markets creates obstacles to this idealized process. Inelastic demand results in infrequent price spikes, which make for a risky investment environment. Involuntary curtailments need to be priced at an estimate of consumer valuation. Price caps that are intended to control for the exercise of market power, when set excessively low, further obviate price formation and result in so-called missing money (Hogan, 2005), (Fabra, 2018).

An alternative process for arriving to scarcity pricing relies on the acknowledgement of the increasingly important role of reserves and reliability in future power systems. The theory relies on an explicit valuation of reserves (Hogan, 2013), and the interaction between the value of reserve and energy in market equilibrium. The theory is in resonance with the evolutions that are taking place in modern power systems that are resulting from the large-scale integration of renewable resources.

As we increasingly integrate renewable resources in power grids, we exert downward pressure on short-run marginal costs (since renewable resources typically do not require costly fuels for their operation) while increasing the overall uncertainty experienced by the system (since renewable resources typically fluctuate beyond human control). This

---

* Corresponding author.
E-mail address: papavasiliou@mail.ntua.gr (A. Papavasiliou).

[1] The information and views set out in this paper by Gilles Bertrand and Alain Marien are those of the authors solely and do not necessarily reflect the official opinion of the CREG.

[2] The notion of scarcity pricing can be defined independently of operating reserve demand curves, see section 2.5 of (Stoft, 2002).

https://doi.org/10.1016/j.jup.2022.101488
Received 7 May 2021; Received in revised form 31 December 2022; Accepted 31 December 2022

increased uncertainty has created a shift in value from energy to reserve capacity provided by fast-moving "flexible" resources[3] that are now needed to cover the short-term variations originating from both loads and renewable generation (Papavasiliou, 2020). Scarcity pricing based on the exclusive trading of energy does not explicitly reflect this shift in value. This is indicated by the retirement and mothballing of flexible capacity that has been witnessed in some European markets in recent years.[4] Scarcity pricing implemented through the introduction of price elasticity in the procurement of reserve capacity accounts for this shift in value by producing more frequent and less pronounced price spikes than scarcity pricing based on the trading of energy alone. Instead of relying solely on curtailment events, scarcity rent can also be extracted from the more frequent reserve scarcity events. The resulting "well-behaved" energy price creates a beneficial economic environment for installed flexible assets and reduces the exposure of consumers to highly volatile prices.

1.1. *European balancing markets, European markets for reserve, and Co-optimization evolutions in the EU*

The short-term trading of electricity in Europe is organized through a sequence of consecutive markets. The day before electricity is delivered, the Euphemia market coupling algorithm optimizes the allocation of cross-border capacities between countries and coordinates the dispatch of assets on a pan-European level. This algorithm has a scheduling window of 24 h and an hourly granularity. Agents can modify their position in intraday auctions. The real-time trading of imbalances between the ex-ante dispatch resulting from the day-ahead and intraday auctions and the real-time operations is handled through the balancing markets.

The real-time trading of electricity in Europe is characterized by the coexistence of Balance Responsible Parties (BRPs) and Balancing Services Providers (BSPs). BSPs are owners of flexible assets that offer upward and downward real-time energy to the balancing market, i.e., the real-time market for energy. BSPs are activated by the system operator in real time in order to restore the frequency of the system, and are compensated at the balancing price. BRPs are responsible for portfolios consisting of uncontrollable assets, residential and industrial loads, and flexible assets.[5] BRPs are subject to imbalance settlements based on their real-time imbalance with respect to a forward market position, at an imbalance price that may depend on the balancing price (see articles 9.1 and 9.2 of (ACER, 2020a,b)). This difference in treatment between the remuneration of BSPs and BRPs is justified according to certain TSOs by a different presumed goal for the two entities: "*The imbalance price incentivizes BRPs to keep and/or restore system balance of their imbalance price area in accordance with the Electricity Balancing Regulation, while the balancing energy price reflects the price of the marginal bid selected in the uncongested area by the activation optimization function of the EU balancing platform*" (ELIA, 2021).

In this paper, we refer to "flexibility" as the ability of a resource to provide frequency restoration reserve. Two types of frequency restoration reserves dominate in Europe: automatic frequency restoration reserve (aFRR) and manual frequency restoration reserve (mFRR). aFRR is dispatched by an automatic controller that responds to system frequency and follows a setpoint that varies every few seconds, while the full activation time of the product amounts to a few minutes. mFRR is dispatched by manual instructions. Its full activation time typically amounts to an imbalance interval, namely 15 min.

Both products will be dispatched in the future by European-wide balancing platforms. These platforms are named PICASSO and MARI, for aFRR and mFRR, respectively. The platforms will essentially implement real-time energy markets operated by the TSOs. They are designed so as to maximize the welfare of the system by dispatching balancing capacity from the BSPs in order to clear the imbalances caused by the BRPs.[6] The MARI platform will be run before the beginning of the imbalance interval and can also be launched during the imbalance interval at the request of a TSO (in case of an unforeseen deviation, e.g., an unplanned outage). The PICASSO platform will be run every 4 s during the imbalance interval. Both platforms are expected to include a zonal model of the underlying network. Thus, the platforms will trade transmission capacity implicitly (even if with an imprecise zonal model) in real time. On the other hand, these platforms are not designed to trade reserve capacity in real time. We will therefore refer to these balancing platforms in the present paper as "energy-only."

European legislation anticipates the value migration that is taking place in electricity markets from energy to reserves. Articles 40–42 of the Electricity Balancing Guideline (EBGL) of the European Commission (European Commission, 2017) are central in moving towards an integration of the European day-ahead market for reserves, which is currently fragmented.[7] Three possibilities are foreseen in articles 40–42 of the EBGL for the allocation of transmission capacity for the trading of reserves: co-optimization (article 40), market-based allocation (article 41), and the allocation of reserves based on economic efficiency analysis (article 42).

Under the co-optimization framework of article 40, day-ahead energy bids are known, the available transmission capacity is known, and the reserve bids are known when the day-ahead market clearing model is solved. This implies that the existing day-ahead energy market model, which is solved by EUPHEMIA, would need to be enhanced with reserve products. This is the target option in legislation, i.e., the only option for which TSOs are obliged to propose a method. If TSOs also propose methods for articles 41 or 42, they can choose among the different articles on a bilateral basis. The method of article 40 has already been approved by the Agency for the Cooperation of Energy Regulators (ACER). Methods for articles 41 and 42 have also been proposed. Regardless of the ultimate choice of TSOs regarding their preferred methodology, the legislation acknowledges the special role of co-optimization as the default option. The linking of energy and reserve bids is key in this market clearing model and applies identically to scarcity pricing, as we discuss next.

1.2. *Scarcity pricing with and without Co-optimization*

Article 40 applies to the day-ahead market, whereas MARI and PICASSO will not implement co-optimization in the foreseeable future. The following example illustrates the implications of co-optimization of energy and reserves in terms of pricing in real time, and in terms of scarcity pricing in particular. Consider the (hypothetical) application of co-optimization on an illustrative single-zone example of a balancing market, e.g., MARI. In this hypothetical example, the following bids would be received in the market. All quantities are reported in MWh. For

---

[3] This flexibility can originate from various resources: demand response, storage, or fast-moving thermal units (such as combined cycle gas turbines).

[4] Although the ongoing gas crisis is likely to affect the immediate evolution of the European electricity market, scarcity pricing based on operating reserve demand curves recedes when energy prices alone generate price signals that can support investment in flexible resources. It can thus be considered as a no-regret measure which simply does not interfere with the energy market when energy prices alone are sufficiently high to induce investment in flexible resources.

[5] According to article 18 of the Electricity Balancing Guideline, every BSP balancing energy bid must be attributed to one or more BRPs.

[6] BRPs that include BSPs which have been activated are subject to imbalance adjustments. These imbalance adjustments ensure that the activated BSPs are not be penalized for helping the system.

[7] Whereas the European day-ahead energy market is coupled, existing European day-ahead markets for reserve allow for a trading of reserves before (e.g., Belgium, the Netherlands), or after (e.g., Italy, Spain) the clearing of the day-ahead energy market.

the case of capacities, this unit of measurement should be interpreted as the amount of reserve capacity that is available for 1 h.

- BSP energy supply bids:
  - BSP-A: 10 MWh @ 20 €/MWh
  - BSP-B: 10 MWh @ 50 €/MWh
- Energy demand bids (submitted by TSOs on behalf of BRPs):
  - Price-taking for 15 MWh
- BSP reserve bids[8]:
  - BSP-A: 10 MWh @ 0 €/MWh
  - BSP-B: 10 MWh @ 0 €/MWh
- TSO reserve demand[9]:
  - 10 MW @ 350 €/MWh

The goal is to maximize welfare, which is the sum of the welfare generated by energy trades and reserve trades. Note that the energy and reserve bids of BSP-A and BSP-B are *linked* in the sense that there is an explicit constraint in a co-optimization model that imposes that the sum of energy and reserve provided by each of these BSPs cannot exceed these BSPs' capacities. This co-optimization model is described mathematically in Appendix A2.

The market clearing model matches the cheapest energy resources to the energy demand and keeps the remainder to cover the reserve demand. Concretely, BSP-A is fully accepted and supplies 10 MWh of energy to cover the energy demand bid, and the remaining 5 MWh of demand for energy is covered by BSP-B. BSP-B can now use its remaining capacity, 5 MWh, to cover part of the TSO reserve demand. The price of the reserve is 350 €/MWh since the TSO demand is only partially accepted. In order for BSP-B to be willing to split its 10 MWh between energy and reserves, it must earn the same profit margin from both markets. This implies an energy price of 400 €/MWh, which is the marginal cost of the marginal unit (in this case, BSP-B) plus the reserve price (which we also refer to as a scarcity adder). By contrast, the energy price in an energy-only market would be 50 €/MWh.

Observe how scarcity pricing is at work in this example: the energy price exceeds the marginal cost of the most expensive unit in the market due to the fact that the system is tight. The system is not tight to the point of not being able to satisfy energy demand, but it is tight to the point of not being able to satisfy the demand for reserve fully, and this in itself is sufficient to produce scarcity prices.

The example above produces scarcity pricing through an explicit multi-product auctioning of energy and reserves. As we discuss in the sequel and further justify in past work (Papavasiliou, 2020), an accurate valuation of reserves based on highly dynamic and highly uncertain conditions requires implementing the same procedure in real time. However, European balancing platforms are "energy-only" platforms that do not trade reserve capacity in real time. This is a remarkable oversight in European electricity market design. It does not simply mean that there is no co-optimization of energy and reserve in real time. Instead, reserve is not even traded in real time. This significant market incompleteness creates serious challenges for the valuation of reserves in European balancing markets (Papavasiliou, 2020).

The difference between not co-optimizing energy and reserve in real time and not trading reserve at all cannot be overstated. Note that scarcity pricing can be implemented even in the absence of co-optimization. The original Texas design of scarcity pricing based on Operating Reserve Demand Curves (ORDC) is a case in point (ERCOT, 2014). The idea in this design is to trade reserve implicitly by computing the price of reserve as the evaluation of the reserve demand curve at the point of remaining reserve capacity, which can be measured based on real-time telemetry. We refer to this as implicit trading of reserve capacity.

To illustrate the concept of implicit trading of reserve, let us revisit our simple example. The first step in reproducing the outcome of the co-optimization is to run the energy-only platform. As in the case of co-optimization, the platform matches 10 MWh of BSP-A and 5 MWh of BSP-B with the 15 MWh of energy demand. One computes an equilibrium price of 50 €/MWh since BSP-B is marginal. Through telemetry, it is possible to then measure the leftover capacity in the system, which equals the 5 MWh of BSP-B. We can then compute the reserve price as the valuation of the TSO demand curve for 5 MWh of reserve capacity, i.e., 350 €/MWh. We then add this "scarcity adder" to the energy price computed by the platform in order to settle BSPs that respond to system needs under conditions of scarcity.

The MARI and PICASSO platforms are expected to be energy-only platforms that do not trade reserve capacity. In this sense, the future integrated European balancing market will likely resemble the Texas real-time market of (ERCOT, 2014), as it will not feature co-optimization of energy and reserve in real time. This does *not* mean that it should not feature trading of reserve capacity in real time either.

1.3. Contribution of our paper: scarcity pricing in the EU without Co-optimization

The contributions of this paper are twofold. First, we describe our proposal for the implementation of scarcity pricing in an EU balancing market framework based on the Belgian experience (Papavasiliou et al., 2019), and we compare it to competing counter-proposals that have emerged recently (Bertrand, 2021). Our proposal is motivated by the considerations developed in the previous paragraphs, and its aim is to transpose the first principles of scarcity pricing to the boundary conditions of European balancing markets as accurately as possible. The proposal relies on three economic principles. The first principle is the law of one price (Cramton and Stoft, 2006), which suggests an alignment between the compensation of balancing energy and imbalances. The second principle is back-propagation, which suggests that a forward market for reserves should be settled against a real-time market for reserves. The third principle is coherent pricing in multi-product auctions, which suggests a coupling of balancing energy and balancing capacity prices. This market design proposal is advocated by the Belgian regulatory authority for energy (CREG) in their study on the implementation of a scarcity pricing mechanism in Belgium (CREG, 2021).

Secondly, the paper provides additional clarifications to specific concerns raised by Belgian stakeholders related to the implementation of scarcity pricing without co-optimization. Upon the request of the CREG, the Belgian transmission system operator (ELIA) has examined the market design proposal of (Papavasiliou et al., 2019) and has raised a number of concerns related to the proposal. These concerns are developed in detail in a public consultation (ELIA, 2020), along with a counter-proposal set forth by the Belgian TSO as an alternative to the implementation of a real-time market for reserve capacity that is proposed by (Papavasiliou et al., 2019). In response to the public consultation, a number of market stakeholders have also provided input to the process. Without aiming at exhaustively addressing all points that are raised in the public consultation, the present paper addresses issues related to (1) cross-border effects, (2) multiple reserve products, (3) legal facets, and (4) the coexistence of scarcity pricing based on ORDC with capacity markets.

The remainder of the paper is structured as follows. Chapter 2 introduces operating reserve demand curves. The proposal for implementing scarcity pricing in the EU balancing market, as well as counterproposals, are described in chapter 3. A number of stakeholder concerns are addressed in chapters 4 to 7. Chapter 8 concludes.

---

[8] We assume, for the sake of the example, that BSPs assign a zero opportunity cost to their offered balancing capacity. This is not crucial to the key message of the example.

[9] We assume that TSO demands *can* be price-elastic in the balancing platforms, and are thus not required to be price-taking.

2. Operating reserve demand curves

All markets that impose a minimum amount of operating reserve requirements have operating reserve demand curves. Thus, all European markets (including Belgium) *already* have operating reserve demand curves. Inelastic requirements for reserve, whether they are dynamic or static, are particular forms of ORDCs: ones in which the willingness to pay of the system operator is very high, and the quantity requested is equal to the reserve requirement of the TSO. The significant distinction is not whether a specific market has an ORDC but whether (i) the pricing of reserve and energy (both in real time, as well as in the day ahead) is consistent with the ORDCs, and (ii) whether these demand curves are consistent with the economic value of reliability that reserve delivers to the system. On the latter issue, the current gas crisis has shown the limits of an inelastic reserve requirement. The urgent reduction of contracted fast reserve in Belgium in order to avoid the risk of extreme procurement cost (ELIA, 2022) is an example of possibly misstating the reliability-cost tradeoff contemplated by TSOs. Until consumers can be mobilized sufficiently to reveal their preferences for reliability, the calibration of these "administrative" demand curves will remain a challenging aspect of any electricity market design. Ignoring the problem does not make it go away.

Two principal categories of ORDCs are used in practice (NYISO, 2019). The first is based on "steps." A special case is the example used in section 1 here and inelastic reserve requirements. A "stepped" curve is a demand curve expressed by a finite collection of price-quantity pairs. Each step has a lower valuation than the preceding step and reflects the decreasing valuation of the TSO for increments of reserve capacity. In the example employed in section 1, there is one step, its valuation is 350 €/MWh, and its quantity is 10 MWh. The second category of ORDCs encountered in practice is based on loss of load probability (LOLP) and value of lost load (VOLL). An ORDC that depends on LOLP and VOLL is roughly expressed as (VOLL-MC)*LOLP(R), where LOLP(R) expresses the loss of load probability of the system given that the system carries an instantaneous amount of reserve R, and MC is a proxy of the marginal cost of the marginal resource in the system. The formula is derived from first principles, in particular, an abstraction of economic dispatch as a two-stage stochastic program, in the appendix of (Hogan, 2013). The formula admits generalizations related to multiple reserve products, the substitutability of reserve products, minimum contingency levels, day-ahead versus real-time variations (Hogan and Pope, 2019), geographically nested reserve requirements, and several other attributes. Note that this second category of ORDCs also produces well-behaved decreasing valuations. Concretely, when the system is tight (i.e., R is low), the valuation for reserve is high, and vice versa. These relationships reflect the fundamental economic reality that the closer the system is to scarcity, the more valuable the contribution of reserve towards preserving system security and keeping the loss of load probability in check, which holds whether we find ourselves in the day ahead or in real time. The value of reserve capacity under tight conditions does not vanish in real time because the system walked out of a tight situation unscathed. The following balancing interval succeeds every balancing market interval. Implementing a real-time market for reserve capacity acknowledges that reliability also has value in real time.

A comprehensive overview of the current status and foreseen evolutions of ORDCs in US and EU markets is provided in (NYISO, 2019), with certain elements summarized in Table 1. Among the US ISOs that implement stepped ORDCs, one currently finds ISO-NE, MISO, SPP and CAISO. ERCOT was the first market to implement an ORDC based on LOLP, and PJM recently followed suit, following a recent FERC order in May 2020. Among the ISOs that use stepped curves,[10] ISO-NE and MISO have received recommendations by their market monitor to transition to ORDCs based on LOLP and VOLL. All aforementioned US ISOs implement a real-time market for reserve capacity.

An ORDC of the second type (i.e., based on LOLP and VOLL) has been implemented in the UK (Flamm and Scott, 2014) and has also been proposed for implementation in Belgium (Papavasiliou et al., 2019), Ireland and Poland. The calibration of the Belgian ORDC is currently being analyzed by the authors on the basis of a production simulation model (Cartuyvels and Papavasiliou, 2022). The model explicitly trades off the fixed short-term operating cost of carrying more reserve in the system (e.g., due to startup and minimum load costs, or higher fuel costs of certain reserves) against the additional security that is afforded to the system due to reduced loss of load or costly emergency measures under stressed conditions. The focus is on an appropriate choice of VOLL, as well as assumptions about the statistical parameters of system imbalance[11] which affect the shape of the ORDC. Similar questions have been analyzed in the past by the Belgian TSO with input from the authors (ELIA, 2018). Comparable investigations are reported in the literature regarding the Texas market (Zarnikau et al., 2020).

3. BRPs, BSPs, and transposing the theory of scarcity pricing based on ORDCs to European balancing markets

As we mention in section 1 of the paper, European market design distinguishes between BRPs and BSPs. Although the two entities differ fundamentally in terms of market operation (since the latter can be counted on for providing reserve in order to balance out the imbalances caused by the former), their distinct role in system operation has also led to a distinct treatment in the balancing market. BRPs face imbalance prices, whereas BSPs face balancing prices, despite the fact that one is supplying energy to cover the imbalances of the other. This naturally leads to arbitrage opportunities, especially since BSPs have the freedom to choose whether they wish to self-dispatch their resources or offer them to the balancing market. The distinction between balancing prices and imbalance settlement also leads to confusion as to how scarcity pricing should be implemented. This confusion is compounded by the absence of a real-time market for reserve capacity in Europe.

The question that arises in concrete terms can be illustrated in reference to the example of section 1.2. As we discuss in section 1.2, we compute an energy price of 400 €/MWh and a reserve price (or scarcity adder) of 350 €/MWh. In a co-optimization framework, there is no ambiguity about what these prices imply for settlement: (i) The energy price is used for settling *energy* deviations from forward setpoints, whether these are due to reserves (think BSPs) responding to system imbalances or the resources (think BRPs) causing the imbalances. (ii) The reserve price is used for settling *reserve* deviations from forward setpoints. In an implicit auctioning framework, what does the implementation of the mechanism entail? Should the reserve price/"adder" (350 €/MWh in our example) apply on top of the balancing energy price of the energy-only platform (50 €/MWh in our example)? Should it apply on top of the imbalance price? Should it apply to the settlement of reserve imbalances?

The proposal presented hereafter is part of the literature which focuses on the modeling and design of balancing markets and imbalance settlement. Recent references include (i) (Matsumoto et al., 2022) who investigate the arbitrage opportunities on imbalance settlement in the Japanese electricity market for a virtual bidder and (ii) (Ocker et al.,

---

[10] It is worth noting that, in US markets, stepped demand curves are not necessarily tied to an estimate of the cost of lost load. That is to say, in the past, not all US ISOs priced reserve scarcity based on the estimated cost of lost load.

[11] The statistical parameters of imbalance distributions are anyways estimated frequently by system operators for the purpose of dynamic dimensioning (De-Vos et al., 2019).

Table 1
Overview of worldwide shortage pricing mechanisms based on ORDC.

Country/System	ORDC based on steps or LOLP/VOLL	Real-time co-optimization of energy and reserve	Real-time market for reserve capacity	Comments
Belgium (Papavasiliou et al., 2019) (CREG, 2021)	LOLP/VOLL	No	Proposed (Papavasiliou et al., 2019) (CREG, 2021)	The mechanism is proposed, not implemented
Poland (PSE, 2019)	LOLP/VOLL	No	No	Planned for implementation by the first half of 2023
UK (Great Britain Department for Business, Energy & Industrial Strategy, 2020)	LOLP/VOLL	No	No	Scarcity adder applied on the imbalance price
Ireland (SEM, 2021), (EirGrid, 2017)	Steps	No	No	
Greece (Papavasiliou, 2021)	LOLP/VOLL	No	Proposed	The mechanism is proposed, not implemented
ERCOT (NYISO, 2019)	LOLP/VOLL	No	Yes	ERCOT is moving forward with the introduction of real-time co-optimization (Harvey, 2020)
PJM (NYISO, 2019)	LOLP/VOLL	Yes (Giacomoni et al., 2018)	Yes	Real-time co-optimization for all reserve products to be introduced in 2022 (Harvey, 2020)
ISO-NE (NYISO, 2019)	Steps	Yes (Zheng & Litvinov, 2006), (ISO-NE, 2013)	Yes	Market monitor recently recommended move to ORDC based on LOLP/VOLL (NYISO, 2019)
MISO (NYISO, 2019)	Steps	Yes (Chen, 2019)	Yes	Market monitor recently recommended move to ORDC based on LOLP/VOLL (NYISO, 2019)
CAISO (NYISO, 2019)	Steps	Yes (Harvey, 2020)	Yes	
SPP (NYISO, 2019)	Steps	Yes (SPP, 2010)	Yes	

2018) and (Kraft et al., 2019) who examine different balancing market designs based on game-theoretical arguments for the European balancing market, with a focus on the German market.

3.1. *Our proposal: a single energy price and a market for real-time reserve capacity*

Our proposal to the Belgian regulator (Papavasiliou, 2020), which we refer to in the present paper as *"First Principles Design"*, has been "all of the above". Concretely, we propose that (i) the value of reserve should be used as a scarcity adder for uplifting the balancing price, (ii) that the imbalance price should be equalized to the balancing price, and that (iii) this same adder should be used for settling reserve imbalances in real time.

The rationale for each element of the proposal relies on first principles of economics. (i) Using a scarcity adder to uplift the balancing price approximates the equilibrium prices of a multi-product auction with both reserve and energy. This adder couples the prices of energy and reserve and eliminates arbitrage opportunities between the energy and reserve market for generators supplying both reserve and energy. The absence of arbitrage opportunities reproduces the incentives that emerge naturally from a multi-product economic equilibrium. (ii) Equalizing the imbalance and balancing price applies the law of one price (Cramton and Stoft, 2006) and eliminates incentives to self-schedule flexible assets. (iii) Using scarcity adders to settle reserve imbalances in real time is based on the fact that the back-propagation of the value of a product or service to forward markets requires that a real-time market is in place for said product or service. The value of the scarcity adder can be derived from the value of lost load and the loss of load probability given the level of remaining reserve capacity in the system, as indicated in the adder formula of section 2.

In order to develop concrete arguments for our proposal, we consider an analytical model of a fringe agent who faces an imbalance price that is potentially different from a balancing energy price, and is called to decide about whether it should bid its capacity to the balancing market or whether it should self-schedule it instead (Papavasiliou, 2020). The analysis is stylized in the sense of considering a risk-neutral fringe agent without fixed (startup or min load) costs for committing reserve, and a unique reserve product. Nevertheless, sound market design choices should stand the test of an idealized setting. Failure to do so suggests fundamental flaws that are likely to be compounded in more realistic settings.

Our analysis is juxtaposed against a number of alternatives in (Papavasiliou and Bertrand, 2021): (i) The *"Belgian Design"*: The existing Belgian balancing market design, which applies an "alpha" component to the balancing price in order to derive an imbalance price which is different from the balancing price whenever the system is very long or very short. Thus, the alpha component is a function of system imbalance.[12] (ii) The *"BRP only design"*: A commonly occurring counter-proposal[13] whereby the scarcity price is added on top of the imbalance price alone.

The Belgian Design is shown in (Papavasiliou and Bertrand, 2021) to have a lesser effect on the back-propagation of scarcity prices to forward markets when alpha is a symmetric function of system imbalance and when system imbalance obeys a symmetric probability distribution. The intuition driving this result is as follows: a symmetric distribution of imbalances, combined with symmetric alpha components, implies an average imbalance price that is equal to the average balancing price. The owner of a flexible asset therefore has an interest in submitting its entire flexible capacity to the balancing market at its true incremental cost, since the balancing market will only activate the capacity when it is profitable for the flexible asset owner. And since the flexible capacity would anyway be bid into the balancing market, there is no opportunity cost involved in bidding the capacity in a forward reserve market. This is a good thing, since the flexible capacity is made available to the TSO in the balancing market. Unfortunately, however, there is no back-propagation of a reserve value to the forward reserve market, since there is no real-time market for reserve capacity.

Regarding the BRP-Only Design, there is an asymmetry between the payoffs of the flexible asset owner from the balancing price versus the imbalance price: the reserve adder creates a payoff from the imbalance price which is, on average, higher than the balancing price. This induces the owner of the flexible asset to self-dispatch and thus take its chances with the imbalance price. This in turn produces an opportunity cost for

---

[12] System imbalance and scarcity, although related, are not the same thing. A system may be very short and still have abundant reserve, even when neutralizing a significant power shortage. This is discussed further, and in relation to the response of the European Commission (European Commission, 2020) to the Belgian electricity market national implementation plan (Federal Public Service Economy, 2019), in (Papavasiliou, 2020).

[13] Paul Giesbertz, "The power market design column – The scarcity of scarcity pricing", available online: https://www.linkedin.com/pulse/power-marketdesign-column-scarcity-pricing-paul-giesbertz/.

the owner of the flexible asset in the forward reserve market, since we assume that any capacity cleared in the forward reserve market cannot be self-dispatched in real time.[14] Unfortunately, however, two adverse side-effects are at play. The first side-effect is that this design induces low-cost flexibility service providers to keep their balancing resources out of the balancing market, and thereby strips the TSO from much-needed flexibility while also resulting in inefficient dispatch in expectation. The second side-effect is that the self-dispatch action of the flexibility provider depresses balancing prices, thereby cancelling the exact effect that scarcity pricing is intended to achieve.

The analysis in (Papavasiliou, 2020) and (Papavasiliou and Bertrand, 2021) concludes through analytical reasoning and agent-based simulation that the First-Principles Design is the only one among the three that induces flexible asset owners to bid their reserve capacity to the balancing market, while also inducing a back-propagation of the real-time value of reserve capacity to forward (e.g., day-ahead) reserve markets.

3.2. *A counter-proposal by ELIA: combining Belgian imbalance adders and scarcity adders on imbalance settlement alone*

As mentioned in section 1, ELIA responded to our market design proposal with an alternative proposal in its public consultation (ELIA, 2020). In addition to the existing “alpha” component mentioned in section 3.1, ELIA proposes introducing an “omega” component on the imbalance price of the BRPs. This component is a variation of the scarcity pricing adder formulas based on LOLP and VOLL discussed in section 2. The only essential difference is that the “omega” formula proposed by ELIA[15] attempts to also account for sustained imbalances by only triggering the adder after one period of high imbalances has transpired. Estimates of the average value of omega compared to the standard scarcity pricing formula for 2018 indicate that the effect of this difference is minor and amounts to a few cents per MWh.

We can build on the analytical framework of (Papavasiliou and Bertrand, 2021) in order to examine the properties of the proposal. Unfortunately, and as intuition suggests, the proposed design inherits the unfavorable properties of the BRP-Only Design under the same set of assumptions as in section 3.2. Concretely, the presence of the alpha component has no material effect on the analysis. Limiting the application of the omega component to imbalance settlement alone induces self-scheduling of flexible resources, and therefore induces these resources to move out of the balancing market, while also exerting a downward pressure on balancing prices and cancelling the intended effects of scarcity pricing. The complete analytical development is presented in chapter 4 of (Bertrand, 2021).

3.3. *An adder on BSPs and BRPs without a real-time market for reserve*

The European Commission has clarified that it is acceptable to add a scarcity adder on top of the imbalance and balancing prices. Concretely, in its decision on the implementation plan of the Belgian Capacity Remuneration Mechanism (European Commission, 2021), the European Commission has requested the Belgian government to consider implementing an adder not only on BRPs but also on BSPs, see *whereas (62)* of (European Commission, 2021). This has motivated some stakeholders to consider a design in which the scarcity adder is applied to BRPs *and* BSPs, but without a real-time market for reserve imbalances. An analysis of the design under the same set of idealized assumptions as in the preceding subsections suggests that this situation is similar to the case where no adder is implemented in the first place. The reason is that BSPs would be incentivized to bid all of their capacity at their marginal cost minus the expected adder because they only receive the adder if they are cleared. Therefore, the final balancing price would remain the same as in the case without an adder. The absence of a real-time market for clearing reserve imbalances, again, impedes the back-propagation of reserve prices to the day ahead.

4. Cross-border interactions

The possibility of a unilateral implementation of scarcity pricing in Belgium raises a number of interesting questions as the European energy-only balancing markets are advancing towards integration. The ELIA public consultation (ELIA, 2020) considers two possible approaches for integrating scarcity prices in cross-border platforms (MARI and PICASSO): (i) applying ex-post adders to the platform prices, or (ii) uplifting the BSP offers by the adders, so as to internalize expected adders as a function of the anticipated level of reserves. The second approach has never been proposed or analyzed by the authors, and it is not clear whether it is institutionally viable or compatible with first principles. The discussion in this section is instead focused on the first option.

4.1. *Interfacing scarcity pricing with the platforms*

Co-optimization and an energy-only market clearing model may produce different dispatch outcomes. Consider the simple example presented in Fig. 1, which provides a basic illustration of the interaction between two markets, where one market implements scarcity pricing (BE) whereas the other does not (FR).

The interpretation of a unilateral implementation of scarcity pricing in a co-optimization model is that the TSO places balancing capacity bids in one market but not the other. The representation of the network constraints in a model that co-optimizes energy and reserves is a question in its own right and relates directly to the implementation of EBGL article 40. Here we are interested in highlighting the rationale of a unilateral implementation of scarcity pricing without co-optimization and can highlight insights without concerning ourselves with this complex question.

With the values indicated in the figure, the outcome of an energy-only market is to activate BSP-BE at 8 MWh so that it can cover the BRP-BE energy demand. The price is 10 €/MWh in both locations since the link is not congested. Instead, a co-optimization model only activates BSP-BE by 5 MWh and saves the remaining capacity for covering a part of the TSO-BE demand for reserve. The remaining 3 MWh of energy demand needed to satisfy BRP-BE are imported from BSP-FR.

Note that the dispatch is different between the two models. The

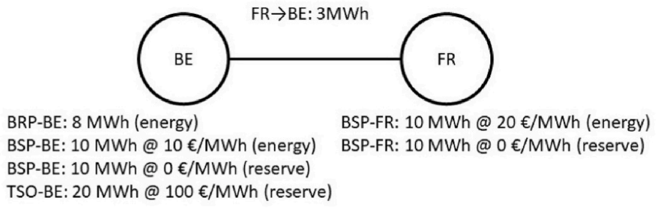

Fig. 1. Illustrative example of the interaction between a market that implements scarcity pricing (BE) and a neighboring market that does not (FR).

---

[14] This is an assumption in our analytical model. Deviations from this assumption may apply in practice. For example, mFRR resources may self-dispatch if they are notified that they are not accepted in the balancing market. The implications of such deviations would need to be analyzed in further detail. Nevertheless, such deviations may further weaken the case for the ability of the design to back-propagate a reserve price to the forward reserve market.

[15] This statement is correct under the condition that the formula is applied correctly. Concretely, the first condition in the top of page 81 of (ELIA, 2020) that defines the omega appears to be incorrect. In our analysis, we have assumed that the intention was for the logical condition to be for omega to be zero if the adder of the previous period is zero (not the value of omega in the previous period). In the form described in (ELIA, 2020), the expression published by ELIA amounts to a constant value of 0 €/MWh.

energy supply of BSP-BE is held back in co-optimization so that the capacity can be used to cover the demand for reserve in BE. There is value added in doing so because the benefit derived by covering more reserve demand in BE exceeds the additional cost of the supplier in FR. The reserve price in the co-optimization model is 100 €/MWh (since TSO-BE is at the money), and the energy price of BE is 110 €/MWh (since BSP-BE should be indifferent between reserves and energy). The energy price in FR is 20 €/MWh.

If the BE zone implemented scarcity pricing unilaterally, it would seek to price energy and reserve to support the dispatch of the energy-only model. Although it is guaranteed by duality theory that such prices exist for the optimal solution of the co-optimization model, it is not guaranteed that they also exist for the optimal solution of an energy-only model. On the other hand, in sufficiently simple settings, a co-optimization model can respect merit order within any given zone: even if units supply reserve, they are dispatched for energy in merit order. The intuition driving this result is that the most expensive units cover reserves; therefore, one requires more complex interactions (e.g., related to maximum reserve limits of individual units) to disrupt merit-order dispatch (NYISO, 2019).

This observation can rationalize the idea of including scarcity adders on top of the platform price. Even if we fix the dispatch of resources to the solution of the energy-only platform, we can respect the no-arbitrage conditions of the resources within the zone that implements scarcity pricing so long as reserve and energy prices are aligned. Resources out of the system are guaranteed to respect their no-arbitrage conditions by balancing platform design. Concretely, BSP-FR does not require a scarcity adder to deliver energy to the platform: the energy price of the platform for FR is already sufficient for this purpose.

4.2. Dumping

Dumping refers to a practice whereby producers bid below marginal cost due to distortions in market design. Dumping has been raised as a concern in the context of a unilateral implementation of scarcity pricing in Belgium since Belgium will soon connect to the European cross-border balancing platforms MARI and PICASSO. Specifically, the design should incentivize BSPs to bid at their marginal cost and BRPs to react if their marginal cost is cheaper than the balancing price, as would be the case in a design without any adder. This result can also be achieved by the First-Principles Design proposed by the authors because, on the one hand, BSPs are incentivized to make their flexibility fully available in the balancing market, and on the other hand, BSPs are exposed to the adder no matter if they produce or not. BSPs, therefore, are not incentivized to resort to dumping in order to be remunerated by the adder. Nevertheless, this property does not hold in the case of the BRP-Only design. The reason is that, in the BRP-Only Design, BRPs are incentivized to react below their marginal cost to receive the adder. The same problem arises if the adder is placed on BSPs and BRPs only (see section 3.3) because BSPs are incentivized to bid below their marginal cost to increase the likelihood of being activated and, thus, receiving the adder.

4.3. TSO cash flows

The interpretation of scarcity pricing based on ORDC as a real-time auction also suggests how the mechanism can be financed. The rationale of forward (e.g., day-ahead) reserve auctions is that the TSO procures reliability, a public good, on behalf of the entire market. A real-time (even if implicit) reserve auction would follow the same principle. Note that this is an auction for reserve "imbalances". Thus, even though the implicated cash flows apply to lower quantities, can cancel each other out, and may therefore be relatively minor compared to forward markets where the bulk of trading takes place, the mere presence of the real-time market can drive price formation in forward markets due to back-propagation.

Apart from the financing of reserve imbalances, there remains the question of the financing of the adder with respect to energy imbalances. The concern is that the EU balancing platforms are not foreseen to cover this cash flow. This concern may be mitigated by the fact that the balancing market is expected to be approximately energy neutral, meaning that the balancing activations of one country to cover the imbalances of another country should cancel each other out on average.

5. Multiple reserve products

The scarcity pricing market design proposal developed by (Papavasiliou et al., 2019) *assumes* a unique energy price in order to higher-level questions related to virtual trading, co-optimization of energy and reserves in the day ahead, and co-optimization of energy and reserves in real time, an assumption in the analysis. The fact that this is an assumption of the analysis should not be confused with a claim that cross-product pricing (i.e., equalizing the prices of MARI and PICASSO) is a prerequisite for the implementation of scarcity pricing. The implication of different balancing energy prices emerging from the platforms is a complex question of the EU balancing market design that will need to be analyzed in further detail moving forward.

The system operator may approach the valuation of reserves in diverse ways. For example, the fact that the EU balancing market operates two separate platforms for balancing energy in real time (MARI and PICASSO) may lead one to conclude that the valuation of mFRR and aFRR capacity are independent. One would then define two functions, $V^{mFRR}(d^{mFRR})$ and $V^{aFRR}(d^{aFRR})$. We can refer to this model of TSO valuation as *independent valuations*. However, certain system operators size reserves using a probabilistic methodology that first defines an FRR target and splits this FRR capacity between mFRR and aFRR. The split must obey rules related to the minimum capacity that the system should carry in aFRR capacity. Thus, we may instead assume that the system operator is employing two functions, $V^{FRR}(d^{aFRR}+d^{mFRR})$ and $V^{aFRR}(d^{aFRR})$. We can refer to this model of TSO valuation as *interdependent valuations* (the terms independent and interdependent valuations used in the context of scarcity pricing are due to William Hogan).

The implementation of scarcity pricing in the presence of multiple reserve products is best understood when considering a co-optimization model and one-way substitutability. A model that co-optimizes energy and reserves and includes multiple reserve products with interdependent valuations accounts for the fact that "fast" reserve products (e.g., aFRR) can contribute towards the total reserve needs of the system (i.e., for FRR). Concretely, a model with interdependent valuations can be formulated to produce an equilibrium price for energy, fast reserves, and total reserves. This model then implies payments for those resources which supply energy, those resources which can supply "fast" reserve (e.g., aFRR), and those resources which can supply "slow" reserve (e.g., mFRR), with the latter being resources that are flexible but not rapid enough to supply the highest-quality reserve. All of these prices are coherent in that they support dispatch outcomes that respect the various constraints of the co-optimization model. These constraints include ramp rates, the one-way substitutability of different reserve products,[16] and the maximum capacity limits of individual generators.[17] The prices are also coherent with the valuation of the TSO for different types of reserve capacity, i.e., the demand functions $V^{FRR}(d^{aFRR}+d^{mFRR})$ and $V^{aFRR}(d^{aFRR})$ mentioned in the previous paragraph.

A notable challenge in a market design that trades reserve capacity implicitly is to derive adder formulas in advance. The reserve prices are directly implied by the demand functions of the TSO, and imply real-time prices for "fast" and "slow" reserve capacity (see slide 35 of (ERCOT, 2014)). The original ERCOT design also derives energy adders

---

[16] This refers to the fact that aFRR can contribute to FRR.

[17] This is referred to in European jargon as "linking of bids", and reflects the fact that the total capacity of a resource imposes a limit on the amount of energy and total reserve capacity that the resource can offer.

(see slides 44 and 46–50 of (ERCOT, 2014)), which have served as an inspiration for the Belgian design proposal (Papavasiliou et al., 2019). A simplified co-optimization model of energy and reserves with interdependent valuations can be used to show that, under simplifying assumptions (e.g., in the absence of individual unit reserve constraints), the equilibrium energy price becomes equal to the marginal cost of the marginal unit plus a scarcity adder defined as in slide 35 of (ERCOT, 2014). This reasoning can motivate applying the adder mentioned above to the MARI and PICASSO prices. On the other hand, it is worth noting that the identity mentioned above that relates to the dual multipliers of the energy and reserve balance constraints may not apply in more complex settings that include, for instance, individual resource limits on reserve capacity (NYISO, 2019). This topic, therefore, merits further analysis.

6. Legal facets

Concerning the possibility of an implementation of a scarcity pricing mechanism in a European Member State, the current European legislation allows a unilateral implementation of such a mechanism, either in accordance with article 20 of Regulation (EU) 2019/943 (European Parliament, 2019) or in accordance with article 18 of the Electricity Balancing Guideline (European Commission, 2017). Given the foreseen implementation of several balancing platforms before or simultaneously with a scarcity pricing mechanism in Belgium, the question of providing adequate incentives to market players by the proposed mechanism is of utmost importance.

Article 44(3) of the Balancing Guideline allows the application of a scarcity adder (to settle the procurement costs of balancing capacity) on the imbalance price charged on BRPs. This is confirmed by ACER Decision 18/2020 on imbalance settlement harmonization, where article 9.6 (a) allows a scarcity component to be used in nationally defined scarcity situations.

The application of a scarcity adder to the settlement of BSPs has been challenged in market stakeholder discussions (ELIA, 2020). The debated article is article 20(3).c of Regulation (EU) 943/2019, which reads: “When addressing resource adequacy concerns, the Member States shall in particular take into account the principles set out in Article 3 and shall consider: (c) *introducing a shortage pricing function for balancing energy as referred to in Article 44(3) of Regulation (EU) 2017/2195”;* Specifically, the disagreement is on the interpretation of the reference to article 44(3) *of Regulation (EU) 2017/2195.* On the one side, the interpretation of the CREG is that the reference relates to the concept of a shortage pricing function that is mentioned in Article 44(3). According to the CREG, article 20(3).c should therefore be understood as allowing the application of a shortage pricing function on BSPs. On the other side, the interpretation of ELIA is that the reference relates to the mechanism of applying a shortage pricing function to BRPs. According to ELIA, the article implies that a shortage pricing function can only be applied to BRPs. When analyzing the two interpretations, the following points should be noted. Firstly, the interpretation of ELIA results in a contradiction with article 20(3).c itself. Indeed, the article unambiguously mentions that the shortage pricing function can be introduced for balancing energy (which is clearly linked to BSPs in Article 2 of Regulation (EU) 943/2019). Secondly, ACER Decision January 2020 (ACER, 2020a,b) on balancing energy (which is applicable to BSPs) confirms the interpretation of the CREG. Indeed, it indicates in article 1(4) that this pricing methodology (applicable to BSPs) is without prejudice to the introduction of a shortage pricing function for balancing energy (thus applicable to BSPs), as referred to in article 20(3).c of Regulation (EU) 943/2019. Third, the application of an energy adder to both BRPs and BSPs is considered in the European Commission opinion on the Belgian implementation plan (European Commission, 2020), when indicating that: “*The Commission, however, invites Belgium to consider whether the scarcity pricing function should apply not only to BRPs but also to balancing service providers (BSPs)*”.

ELIA has further raised state aid concerns regarding the implementation of scarcity pricing (ELIA, 2020). For state aid to exist, four cumulative conditions must be fulfilled (European Commission, 2018a, 2018b): the measure must (i) be funded through State resources and must be imputable to the State, (ii) confer an economic and selective advantage to certain undertakings, (iii) distort or threaten to distort competition, and (iv) be liable to affect trade between Member States. The proposed scarcity pricing mechanism benefits from a natural (through the balancing platforms) possibility of cross-border participation and the absence of any selection between market players. The mechanism is technology neutral, uses existing balancing mechanisms, is open to producers, storage, and consumers, and does not require long lead times for selection or activation. Therefore, condition (ii) should not be fulfilled. Ultimately, it will be up to the Belgian state to decide whether to notify the European Commission of the measure. In three recent decisions of the European Commission on state aid related to the implementation of capacity mechanisms in Poland, the UK, and Belgium, the Commission supports or considers the implementation or improvement of scarcity pricing mechanisms in the scope of the implementation of article 20 of Regulation (2019)/943 (European Commission, 2018a, 2018b), (European Commission, 2019), (European Commission, 2021).

Finally, it is worth mentioning that the process of adopting terms and conditions related to balancing, which includes the settlement of the balancing service providers and of balancing responsible parties, is foreseen by article 18 of the Electricity Balancing Guideline (European Commission, 2017). Combined with article 5.4 (c) of the same guideline, this requires the approval by each (individual) regulatory authority of each concerned Member State on a case-by-case basis.

7. Coexistence with capacity markets

It is misleading to claim that one must choose between scarcity pricing and capacity markets. Scarcity pricing can and does coexist with capacity markets in some US designs, including ISO-NE and PJM (NYISO, 2019). On the other hand, precedence is relevant. The fact that scarcity pricing diminishes missing money problems and, therefore, the scope of capacity markets should not be confused with implying that scarcity pricing cannot coexist with capacity markets.

Both capacity markets and scarcity pricing have perceived advantages and disadvantages. A capacity market provides a clear cash flow for investors and mitigates revenue certainty concerns of risk-averse investors. However, capacity mechanisms can be prone to a myriad of design challenges and can result in the over-procurement of costly capacity. Since the procurement cost is passed on to consumers, and TSOs are not required to cover this procurement cost, these mechanisms are preferred by some market stakeholders. Scarcity pricing can avoid the long-term commitment to possibly excessive capacity procurement, but difficulties may emerge in the alignment of the mechanism with missing markets in European balancing markets (notably the absence of a real-time market for reserve capacity).

Both methods also share the disadvantage of relying on administrative parametrization (Cramton, 2005). Scarcity pricing based on ORDC requires estimating the operating reserve demand curve, hence VOLL, and the distribution of imbalances. Capacity markets require an estimation of the capacity demand curves (including VOLL) and a myriad of other challenging design parameters (eligibility, frequency, and look-ahead of auctions, derating factors, and many other non-obvious administrative choices).

Both methods attempt to compensate for a major endemic challenge in electricity markets: the absence of short-term demand-side price elasticity. Scarcity pricing and capacity markets might be temporary crutches that will cease to be useful when electricity markets can mobilize responsive demand at a sufficiently wide scale.

7.1. Double payments

The basic analysis by (Boiteux, 1960) can be generalized to assess the question of double payments when the two mechanisms coexist. Concretely, we can analyze a long-term investment equilibrium where scarcity pricing based on ORDC coexists with a capacity remuneration mechanism. This analysis leads to key observations: (i) Double payments do not occur in a long-term perfect competition equilibrium of greenfield investment, in the sense that generators only recover enough short-term revenues from the energy, reserve, and capacity market to cover their investment costs. Should they recover more, they would be undercut by competitors. (ii) This does not mean that an energy-only market to which we add a capacity mechanism has an investment outcome unaffected by introducing scarcity pricing based on ORDC. The new investment outcome depends on appropriately calibrating the capacity market demand curve. (iii) When capacity mechanisms coexist with scarcity pricing, capacity market revenues may "take up space" in the revenue streams required to cover investment costs of new technologies. This phenomenon may suffocate their energy revenues and suppress their investment in technologies that are unable or ineligible to trade in a capacity market (e.g., certain types of demand response).

7.2. ERCOT 2021 outage

ERCOT does not implement a capacity market, and one may incorrectly conclude that the recent forced outages in Texas resulted from scarcity pricing. The Texas blackouts were driven by disruptions in the gas network and several other extreme weather-related factors. To a large extent, the disruptions were a problem of resilience: even if adequate capacity exists in the system, it is rendered useless if the underlying gas network cannot deliver fuel to thermal units (Cramton, 2021). On the other hand, scarcity pricing provides incentives for investing in demand response (Cramton, 2021). It is unclear whether Texan ratepayers would have been willing to insure against such extreme conditions by investing in the required winterization technology to ride through such extreme weather events. Even if this is the case, it does not invalidate the beneficial role of scarcity pricing.

During forced outages, an energy-only design results in price spikes, and scarcity adders become zero (Hogan, 2013). This situation is also explicit in the Belgian scarcity price formulas, where the adders are equal to zero whenever the price is set by forced outages (Papavasiliou et al., 2019), (ELIA, 2020). A critical discussion about "circuit breakers" is emerging following the Texas postmortem. Preliminary discussions in the market design community focus on adapting circuit breakers to different seasons (Cramton, 2021). Even if a particular system exhibits historical demand peaks in the summer, changing climate patterns may shift this stress unpredictably.

7.3. Natural gas crisis

Over the past months, the electricity market has demonstrated highly unpredictable behavior. The recovery from COVID and several disruptions in gas, coal, and other energy markets, combined with a slowdown in investment in energy production infrastructure, led to surges in various energy prices, including natural gas, coal, and oil. CO2 emissions prices followed suit due to a pickup in CO2-intensive coal consumption. The end effect on electricity prices has been unprecedented, with wholesale electricity prices in the hundreds of euros per MWh becoming common for extended periods. The financial repercussions have been massive for electricity retailers, while certain technologies on the supply side of the market are enjoying comfortable profit margins.

The recent invasion of Russia in Ukraine has introduced further stress on the system. As the market is attempting to absorb the likelihood of supply disruptions of natural gas due to possible physical damages on the Ukrainian gas network or sanctions on Russian energy suppliers, the end effect is a sustained high in natural gas and oil prices. Electricity prices naturally follow suit. The political reactions have been swift, with the US deciding to ban imports of Russian natural gas and the European Union contemplating a price cap on natural gas and electricity prices.

Although scarcity pricing appears as an afterthought in this economic environment of sustained high prices, this environment of massive uncertainty underscores the adaptive nature of the mechanism. When profit margins are comfortable, scarcity adders recede, and variable costs drive equilibrium energy prices. When the system becomes tight, non-zero adders emerge and send an attractive signal to investors at a time when this signal is needed most. This effect should be contrasted to alternative designs that commit a fixed payment and early capacity procurement, regardless of future system conditions.

8. Conclusions

This paper revisits several unique market design challenges that emerge due to the adaptation of scarcity pricing based on an operating reserve demand curve to European balancing markets. It uses co-optimization as a disciplined framework for reasoning about these challenges even though co-optimization is not a prerequisite for implementing scarcity pricing, as shown by the ERCOT example. The resulting insights address questions that have emerged from the public consultation of the Belgian TSO (ELIA, 2020) regarding the scarcity pricing proposal of (Papavasiliou et al., 2019) and the feedback of the market stakeholders to the public consultation.

Our initial proposal for implementing scarcity pricing in European balancing markets was based on three points (Papavasiliou, 2020): (i) Introduce a real-time market for balancing capacity to ensure the back-propagation of reserve prices to forward markets. (ii) Introduce scarcity adders based on ORDC to the balancing price. (iii) Equalize balancing and imbalance prices to respect the law of one price. ELIA's counterproposal relies on adding an "omega" component based on an ORDC to supplement the imbalance settlement of BRPs on top of the "alpha" component. Our analysis suggests that this measure would incentivize self-scheduling and induce downward pressure on the balancing price, thereby cancelling the intended effect of scarcity pricing.

We further discuss specific concerns of the stakeholders relative to (1) cross-border effects, (2) multiple reserve products, (3) legal facets, and (4) the coexistence of scarcity pricing and capacity markets. We can summarize these as follows.

1. Scarcity prices can be applied as adders ex-post to the energy-only pan-European balancing platforms MARI and PICASSO.
2. A model of interdependent TSO valuations for reserve products of different qualities can support the rationale of closed-form adder formulas for aFRR, mFRR, and energy that follow the example of the original ERCOT scarcity pricing design (ERCOT, 2014).
3. Scarcity price adders can be applied to both balancing energy and imbalances (ACER, 2020a,b).
4. Scarcity pricing can coexist with capacity markets, and the fact that it can reduce the scope of capacity markets by contributing towards mitigating the missing money problem should not be confused as implying that scarcity pricing is incompatible with capacity markets.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors are grateful for helpful discussions with Marijn Maenhoudt, Andreas Tirez, William Hogan, and Susan Pope. The work has received funding from the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (grant agreement No. 850540).

Appendix

A1. Glossary

In this section, we summarize terminology used throughout the paper by drawing correspondences between how this terminology is used in the context of US and EU market design.

EU market terminology	Proxy in US market terminology (if any)	Definition
Balancing		Act of resolving real-time energy imbalances
Balancing market	Real-time energy market	Market for resolving real-time imbalances
Balancing energy	Real-time energy	Activated capacity used to cover real-time imbalances
Balancing capacity	Reserve	Unused headroom in a generating unit that can be used to balance the system in real time. Balancing capacity is not traded in balancing energy markets but in balancing capacity markets.
Automatic frequency restoration reserve (aFRR)	Automatic generation control	Frequency restoration reserve driven by an automatic controller
Manual frequency restoration reserve (mFRR)	Contingency reserve	Frequency restoration reserve manually activated, e.g., to cope with contingencies (full activation time: 15 min)
Balancing Service provider (BSP)		Supplier of balancing energy and balancing capacity
Balance Responsible parties (BRP)		Entity that strives to keep its portfolio balanced but may cause real-time energy imbalances
Balancing price		Price received by the BSPs for the activation of balancing energy
Imbalance price		Price paid by the BRPs for the imbalances that they cause in the system
MARI		Pan-European platform for the activation of mFRR
PICASSO		Pan-European platform for the activation of aFRR
Energy-only market		Market that only trades energy and not reserve/balancing capacity

A2. A Balancing Market Model Based on Co-Optimization

This section provides a stylized co-optimization model of energy and reserves. This model allows us to define various terms introduced in the paper unambiguously. It further allows us to justify our economic reasoning based on a well-defined underlying model.

The notation presented hereunder is used to characterize the European balancing market as proposed by (Papavasiliou et al., 2019). The problem maximizes the benefit of the TSO for holding reserve minus the cost of covering imbalances given BSP bids that are submitted at a given marginal cost (C) and for volume (P), given BRP real-time imbalances (IMB) and given the marginal valuation for reserve by the TSO ($ORDC(\cdot)$). The BSP capacity is split between energy (p) and reserve (r) (or balancing energy and balancing capacity in EU terminology), and the duality conditions of the optimization problem unambiguously characterize the remuneration of both products. Balancing energy is remunerated at the energy price, which is the dual variable related to the market clearing constraint for energy (λ). Balancing capacity is remunerated at the reserve price, which is the dual variable related to the market clearing constraint for reserve (λ^R). Note that constraint (4) results in a coupling of the energy and reserve price and that dr represents the remaining balancing capacity in the system.

$$\max \int_0^{dr} ORDC(x)dx - \sum_{i \in BSP} C_i p_i \tag{1}$$

$$s.t. \sum_{i \in BSP} p_i = \sum_{i \in BRP} IMB_i \ (\lambda) \tag{2}$$

$$\sum_{i \in BSP} r_i = dr \ \left(\lambda^R\right) \tag{3}$$

$$p_i + r_i \le P_i, i \in BSP \ (\mu_i) \tag{4}$$

$$p, r, dr \ge 0$$

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

$$0 \le r_i \perp \mu_i - \lambda^R \ge 0 \tag{5}$$

$$0 \le p_i \perp C_i + \mu_i - \lambda \ge 0 \tag{6}$$

$$0 \le dr \perp \lambda^R - ORDC(dr) \ge 0 \tag{7}$$

Constraint (7) links the reserve price to the value of the ORDC at the level of remaining balancing capacity in the system. Let us refer to a marginal generator as a generator that supplies both energy and reserve at the optimal solution, and let us index this generator by g. Constraints (5) and (6) show that the scarcity rent for a marginal generator g is equal to μ_g and is further equal to the reserve price. In this stylized model without ramp constraints, μ_g is equal to λ^R, and the energy price is then equal to the marginal cost of the marginal generator C_g supplemented by the reserve price λ^R. These conditions ensure that a marginal generator is indifferent between supplying reserve or energy.

Without co-optimization, it is possible to induce scarcity pricing through an ORDC, as in the original ERCOT energy-only market in Texas (Hogan and Pope, 2020). Coherent price signals are approximated by supplementing the energy price of the energy-only market with a scarcity adder component derived from the ORDC at the level of the available balancing capacity in the system. This approach uses an implicit ORDC adder in the agents' remuneration in contrast to an explicit ORDC in the market-clearing, as displayed in the objective function (1). The obtained prices approximate the equilibrium prices of a co-optimization model without binding ramp constraints if the energy dispatch of the energy-only market is identical to the one of the co-optimized market. The argument is only valid for markets trading a single reserve product composed of a single zone with a zero-cost function for supplying reserve.

A3. European Balancing Market

In practice, balancing markets in the EU differ from the idealized co-optimization presented in the previous section with respect to several points. This section provides a quantitative model of the European balancing market. This model can be described as follows:

$$\max -\sum_{i\in BSP} C_i p_i \tag{8}$$

$$s.t. \sum_{i\in BSP} p_i = \sum_{i\in BRP} IMB_i \ (\lambda) \tag{9}$$

$$p_i \le P_i \ (\mu_i) \tag{10}$$

$$p \ge 0$$

First, TSOs do not value balancing capacity in real time and run an "energy-only" optimization problem. This absence of a market for real-time balancing capacity raises issues for the forward reservation of balancing capacity, as already pointed out in (Papavasiliou and Bertrand, 2021).

Second, the law of one price is not respected for the trading of balancing energy. BSPs' balancing energy (on the left-hand side of constraint (9)) is remunerated at the balancing price, whereas BRPs (on the right-hand side of constraint (9)) face an imbalance price that is based on the balancing price but may differ from it. In Belgium, an "alpha" component was introduced on top of the balancing price. In the view of the TSO, this would incentivize BRPs to keep their imbalance low. ELIA claims that the difference in pricing between imbalances caused by BRPs and balancing energy caused by BSPs can be justified by the different goals of the two products:

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

In practice, flexible assets belonging to a BRP portfolio can be moved freely between active imbalance and participation in the balancing market, giving rise to arbitrage opportunities. The balancing price is currently "paid-as-bid" in Belgium for aFRR, and "paid-as-cleared" for mFRR, and the pan-European balancing platforms MARI and PICASSO are expected to be "paid-as-cleared."

References

ACER, 2020a. Decision N° 01/2020 of the European Union Agency for the Cooperation of Energy Regulators of 24 January 2020 on the Methodology to Determine Prices for the Balancing Energy that Results from the Activation of Balancing Energy Bids. Agency for the Cooperation of Energy Regulators, Ljubljana.

ACER, 2020b. Methodology for the Harmonisation of the Main Features of Imbalance Settlement. Agency for the Cooperation of European Energy Regulators, Ljubljana. annex 1.

Bertrand, G., 2021. *Enhancing the Value of Flexible Generation Capacity through Continuous Intraday Markets and Scarcity Pricing,* Louvain la Neuve: UCLouvain, PhD thesis.

Boiteux, M., 1960. Peak-load pricing. J. Bus. 33 (2), 157–179.

Cartuyvels, J., Papavasiliou, A., 2022. Calibration of operating reserve demand curves using a system operation simulator. IEEE Trans. Power Syst.

Chen, Y., 2019. MISO market design overview and evolution. In: Midcontinent Independent System Operator.

Cramton, P., 2005. A capacity market that makes sense. Electr. J.. vol 18, no 7, pp 43-54.

Cramton, P., 2021. Lessons for Peru from the 2021 Texas electricity crisis. Utility Dive. https://www.utilitydive.com/news/lessons-from-the-2021-texas-electricity-crisis/596998, 23 March.

Cramton, P., Stoft, S., 2006. The Convergence of Market Designs for Adequate Generating Capacity with Special Attention to the CAISO's Resource Adequacy Problem. MIT CEEPR, Cambridge, MA.

CREG, 2021. Study on the Implementation of a Scarcity Pricing Mechanism in Belgium. Commission for Electricity and Gas Regulation, Brussels.

De-Vos, K., et al., 2019. Dynamic dimensioning approach for operating reserves: proof of concept in Belgium. Energy Pol. 124, 272–285.

EirGrid, 2017. Training Modules, Chapter 10: Administered Scarcity Pricing and Reserve Scarcity Pricing. EirGrid Group, Dublin.

ELIA, 2018. Study Report on Scarcity Pricing in the Context of the 2018 Discretionary Incentives. ELIA, Brussels.

ELIA, 2020. Final Report on Elia's Findings Regarding the Design of a Scarcity Pricing Mechanism for Implementation in Belgium. ELIA, Brussels.

ELIA, 2021. Modification of the Alpha Parameters in the Tariff for Maintaining and Restoring the Residual Balance of Individual Access Responsible Parties. ELIA, Brussels.

ELIA, 2022. Consultation Report - LFC Block Operational Agreement. ELIA, Brussels.

ERCOT, 2014. ERCOT Market Training: Purpose of ORDC, Methodology for Implementing ORDC, Settlement Impacts for ORDC. Electric Reliability Council of Texas, TX.

European Commission, 2017. Commission Regulation (EU) 2017/2195 of 23 November 2017 Establishing a Guideline on Electricity Balancing. European Commission, Brussels.

European Commission, 2018a. State Aid No. SA.46100 (2017/N) – Poland – Planned Polish Capacity Mechanism. European Commission, Brussels.

European Commission, 2018b. State Aid SA.42011 (2017/N) – Italy – Italian Capacity Mechanism. European Commission, Brussels.

European Commission, 2019. Commission Decision of 24.10.2019 on the Aid Scheme SA.35980 - 2019/C United Kingdom – Electricity Market Reform: Capacity Mechanism. European Commission, Brussels.

European Commission, 2020. Commission Opinion of 30/04/2020 Pursuant to Article 20 (5) of Regulation (EC) No 2019/943 on the Implementation Plan of Belgium. European Commission, Brussels.

European Commission, 2021. Commission Decision of 27.8.2021 on the Aid Scheme SA.54915 - 2020/C (Ex 2019/N) Belgium – Capacity Remuneration Mechanism. European Commission, Brussels.

European Parliament, 2019. Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the Internal Market for Electricity (Recast). Official Journal of the European Union, pp. 1–71.

Fabra, N., 2018. A primer on capacity mechanisms. Energy Econ. 75, 323–335.

Federal Public Service Economy, 2019. *Belgian Electricity Market: Implementation Plan,* Brussels. Federal Public Service Economy.

Flamm, A., Scott, D., 2014. Electricity Balancing Significant Code Review - Final Policy Decision. Office of Gas and Electricity Markets, London.

Giacomoni, K., Patel, K., Velasco, C.M., 2018. Price Formation Education 3. Reserves and Co-Optimization, PA: PJM Interconnection.

Great Britain Department for Business, Energy & Industrial Strategy, 2020. GB Implementation Plan. Department for Business, Energy & Industrial Strategy, London.

Harvey, S., 2020. Scarcity Pricing Background Discussion. California ISO, CA.

Hogan, W., 2005. On an 'energy Only' Electricity Market Design for Resource Adequacy. Center for Business and Government, JFK School of Government, Harvard University, Cambridge, MA.

Hogan, W., 2013. Electricity scarcity pricing through operating reserves. Econ. Energy Environ. Pol. 2 (2), 65–86.

Hogan, W., Pope, S., 2019. PJM Reserve Markets: Operating Reserve Demand Curve Enhancements. Center for Business and Government, JFK School of Government, Harvard University, Cambridge, MA.

Hogan, W., Pope, S., 2020. Real-Time Co-optimization with the ERCOT ORDC. Harvard University, Cambridge, MA.

ISO-NE, 2013. Operating Reserve Training. ISO New England, MA.

Kraft, E., Ocker, F., Keles, D., Fichtner, W., 2019. *On The Impact Of Voluntary Bids In Balancing Reserve Markets.* s.l., 16th International Conference on the European Energy Market.

NYISO, 2019. Ancillary Services Shortage Pricing. New York ISO, NY.

Ocker, F., Ehrhart, K.-M., Belica, M., 2018. Harmonization of the European balancing power auction: a game-theoretical and empirical investigation. Energy Econ. 73, 194–211.

Papavasiliou, A., 2020. Scarcity pricing and the missing European market for real-time reserve capacity. Electr. J. 33 (10).

Papavasiliou, A., 2021. An overview of probabilistic dimensioning and scarcity pricing with a focus on the Greek electricity market. https://www.rae.gr/wp-content/uploads/2021/05/Report-II-ProbabilisticDimensioning-final.pdf technical report.

Papavasiliou, A., Bertrand, G., 2021. Market design options for scarcity pricing in European balancing markets. IEEE Trans. Power Syst.. vol. 36, no. 5, pp. 4410-4419, Sept. 2021.

Papavasiliou, A., Smeers, Y., De-Maere-D'Aertrycke, G., 2019. *Study on the general design of a mechanism for the remuneration of reserves in scarcity situations,* Louvain la Neuve. Center for Operations Research and Econometrics, UCLouvain.

PSE, 2019. KONCEPCJA Zmian Zasad Unkcjonowania Rynku Bilansujacego W Polsce. Polskie Sieci Elektroenergetyczne, Warsaw.

SEM, 2021. Discussion Paper and Call for Evidence on Scarcity Pricing and Demand Response in the SEM. Single Electricity Market Operator, Dublin.

SPP, 2010. *MOPC Workshop Series On Future Markets: Session 1,* AR: Southwest Power Pool.

Stoft, S., 2002. Power System Economics: Designing Markets for Electricity. IEEE press, Piscataway.

Zarnikau, J., Zhu, S., Woo, C., Tsai, C., 2020. Texas's operating reserve demand curve's generation investment incentive. Energy Pol. 137.

Zheng, T., Litvinov, E., 2006. Ex post pricing in the co-optimized energy and reserve market. IEEE Trans. Power Syst. 4, 1528–1538.

Matsumoto, T., Bunn, D., Yamada, Y., 2022. Mitigation of the inefficiency in imbalance settlement designs using day-ahead prices. IEEE Trans. Power Syst.., vol. 37, no. 5, pp. 3333-3345, Sept. 2022.