Svk Project on Scarcity Pricing

Report on Design Principles

Monday, October 31, 2022

Table of Contents

Executive Summary 2

Principles of Scarcity Pricing and Literature Review **Error! Bookmark not defined.**

Principles of Scarcity Pricing [23] 9

Literature Review 12

Specificities of Nordic system operation 15

ORDC Calibration: Specialized Topics 19

Adaptation of Paradigm to the Swedish / Nordic system 24

Data Input 24

Single-Area Model 26

Multi-Area Model 28

ORDC in a Multi-Area Model 30

ORDC in SE4 Only 35

Separate ORDCs in Each Swedish Zone 41

Where to Apply the Adders, Flexibility Incentives and Back-Propagation 46

Back-Propagation and the Law of One Price 46

Quantitative Framework 48

Existing European Balancing Market Design 50

A Disciplined Approximation of Co-Optimization 51

Applying Adders on Imbalances Only 53

Implementation 56

Legal and Institutional Compliance 56

Affected Products 60

Compatibility with CRMs 63

Cross-Border Effects and Pan-European Balancing Platforms 66

Energy-Only Platforms and Co-Optimization Can Produce Different Outcomes [23] 66

Bidding Incentives and Efficiency of Dispatch [6] 68

No adder 68

Applying Adders on Imbalances Only 68

A Disciplined Approximation of Co-Optimization 68

TSO Cash Flows [6] 69

References 70

Appendix A: Notation of Model (1.2) 74

Appendix B: Notation of Model (3.5) 75

Appendix C: Proof of Coexistence between Scarcity Pricing and CRMs [2] 75

Executive Summary

This report documents phase 1 of the study of a scarcity pricing mechanism for the Swedish power system.

Scarcity pricing based on operating reserve demand curves (ORDC) is a market design proposal which aims at introducing the trading of reserves in real-time markets using a price-elastic demand curve for real-time balancing capacity. The aim of the mechanism is to produce real-time price signals in the real-time market which are awarded to both standby unused capacity as well as to balancing energy that is activated during tight system conditions, with the goal of paying flexible assets for performing under stressful system conditions, when the system needs them most. The mechanism thus targets both system flexibility as well as system adequacy, by generating revenues on top of the marginal cost of the marginal unit, referred to as “adders”, which are payable to capacities that are online under tight system conditions. These adders are intended to be lower but more frequent than alternative real-time market designs, and which are thus intended to generate a more reliable long-run investment signal. The mechanism has seen broad application in US markets, either through direct real-time co-optimization of energy and reserves, or though the ex-post application of scarcity adders that depend on the instantaneous loss of load probability and the value of lost load of the system. The mechanism is also discussed or approximately implemented in a number of European member states.

The calibration of the mechanism depends on a number of design choices. These include how we measure real-time scarcity (e.g. before or after activation within an imbalance interval), the value of lost load, and in the case of multiple reserve products the correlation of imbalance increments within an imbalance interval.

The Swedish system is particularly interesting in the context of scarcity pricing. The system experiences system-wide scarcity during the winter, however it also experiences scarcity in the southern part of the country during the summer, when corridors are congested and flexibility cannot be made available in the south, even if the total net demand of the system during those months is not as high as during the winter. These scarcities are confirmed by a proof of concept that we develop for the Swedish system, which consists of models with escalating complexity. (i) We first implement a single-area model without transmission constraints, which reveals system-wide scarcity. (ii) We then propose a multi-area scarcity pricing model with an ORDC in SE4 which is based on an innovative modeling approach that has been developed by NSIDE in the context of the implementation of energy and reserves co-optimization over networks. The multi-area model reveals system-wide scarcity during the winter, but also local scarcity in SE4 during the summer. (iii) We finally test a multi-area model with ORDCs in all four areas. This interestingly eliminates congestion, and reveals only system-wide scarcity, as in the case of the single-area model. (iv) We use this proof of concept to indicate how strategic and disturbance reserve should be considered in the context of the mechanism. In the process of developing these models, we analyze the optimal business rules for pricing balancing energy and real-time reserve imbalances. The accuracy of our proposal for implementing scarcity pricing without requiring MARI to transition to co-optimization will be tested in detail in phase 2 of the project. These business rules are depicted graphically in Figure 1, and outline how scarcity adders can be applied in the absence but also the presence of congestion.

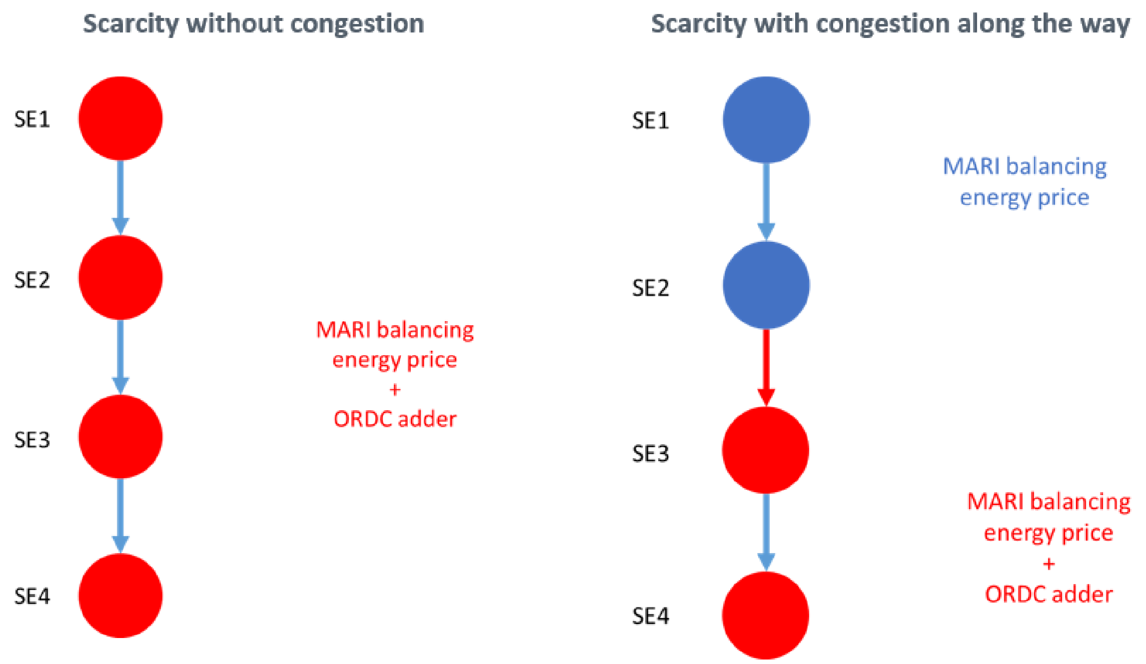

Figure 1: Scarcity pricing in the MARI energy-only platform in the presence of a network: proposal for the Swedish system.

The implementation of scarcity pricing in a way that is compatible with the European balancing market design requires a careful clarification of what scarcity pricing implies for real-time settlements. We focus our analysis on the hypothetical implementation of scarcity pricing to settlements of mFRR activations within MARI, as well as the settlement of capacities that qualify for the response speeds that correspond to mFRR. Although MARI will not conduct an energy and reserves co-optimization in the foreseeable future, it is possible to approximate the outcome of an energy and reserves co-optimization by carefully applying scarcity adders ex-post. We present an analysis of various design alternatives for where these adders should be applied. The analysis is anchored on the economic principles of the law of one price (first stated by Jevons in 1879), arbitrage between energy and balancing capacity, and the back-propagation of real-time prices to forward market prices, and is based on a quantitative model which assumes risk-neutral agents, and zero economic cost for the provision of reserve (only opportunity cost). We analyze a vanilla European balancing market design (with equal balancing prices and imbalance settlement and no market for real-time reserve), a design proposed by Papavasiliou whereby scarcity prices are applied to both balancing prices for BSP settlement as well as imbalance settlements for BRPs, and where the adders are also applied to excess reserve capacity that is available in real time, and an alternative that has been considered by various stakeholders where scarcity adders only apply to imbalance settlement for BRPs. We find that the approach proposed by Papavasiliou preserves the incentive of BSPs to bid their flexible capacity voluntarily into the balancing market while generating a non-zero value for reserve in real time which is back-propagated to forward reserve markets, whereas the proposal of only applying adders to imbalance settlements undermines the willingness of flexible asset owners to bid their capacity voluntarily to the balancing market and results in the back-propagation of a weakened scarcity signal to forward reserve markets. These insights are summarized in Table 1.

	Adder on balancing price?	Adder on imbalance settlement?	Real-time market for reserve?	Back-propagation of real-time value of reserve?	Bid flexibility in balancing market?
"Vanilla" EU balancing market	No	No	No	No	Yes
Disciplined approximation of co-optimization	Yes	Yes	Yes	Yes	Yes
Adder on imbalance settlement only	No	Yes	No	Weak	Not always

Table 1: A summary of the pros and cons of alternative design options for implementing scarcity pricing.

The legal and institutional compliance of the various design alternatives discussed in the previous paragraph are analyzed on the basis of the Clean Energy Package, the Electricity Balancing Guideline, recent ACER decisions, recent European Commission decisions on national implementation plans, as well as recent sector inquiries of the European Commission regarding the implementation of capacity remuneration mechanisms. Arguments are specifically presented as to why the proposal of applying scarcity adders on both BSP and BRP settlement is legally compliant. It is further argued that scarcity pricing is perfectly compatible with the simultaneous implementation of capacity remuneration mechanisms, and the claim that one has to choose between one or the other mechanism is debunked. It is specifically argued that the elimination of missing money through an energy-only market enhanced by scarcity pricing can reduce the value of capacity prices, but does not imply that there are double-payments or that the two mechanisms cannot coexist if deemed necessary. The complexity of extending the scarcity pricing mechanism beyond mFRR remuneration to multiple reserve products is outlined, and reveals more general market design questions of coherent European balancing market design which extend beyond the scope of the current study.

The cross-border properties of the mechanism are analyzed under the various design choices discussed previously, namely a disciplined implementation of scarcity pricing based on co-optimization business pricing rules versus an application of scarcity adders on BRPs only. When embarking on this analysis, it is important to first point out that co-optimization and energy-only pricing can in general produce different dispatch outcomes, which means that there is no guarantee that introducing adders can preserve network equilibrium conditions. This concretely means that an area which implements scarcity pricing can be remunerating its local flexible resources with the balancing price plus an adder, and that neighboring flexible resources in member states which do not implement scarcity pricing are not entitled to such a remuneration. This is nevertheless consistent with the local incentives of flexible resources in member states that do not implement scarcity pricing. Implementing a scarcity adder on BSPs, BRPs and real-time reserve settlement further preserves the incentive of resources to submit their entire flexibility truthfully to balancing platforms, and thus generalizes previous observations about the desirable attributes of the proposal of Papavasiliou to a cross-border setting. It is further argued that TSOs do *not* pay their local resources adders for resolving neighboring imbalances when implementing the mechanism unilaterally, but rather pay adders to the leftover capacity available in their zones.

Context of the Project

The present report is the deliverable of phase 1 of a project titled “Scarcity pricing mechanism for the Swedish / Nordic balancing market”. The project has been commissioned by Svenska Kraftnät in the context of the framework agreement for the provision of expert advisory and analytical support on electricity market design and quantitative analysis.

The project aims to study the possible design of a scarcity pricing mechanism for the Swedish / Nordic balancing market. As part of the redesign of the Nordic balancing markets, the introduction of scarcity pricing is under consideration. The goal of Svenska kraftnät in this project is to gain greater insights into possible models for introducing scarcity pricing in the Swedish/Nordic balancing market.

The specific goals of Svenska kraftnät in this project are to:
i) Develop a proposal for the introduction of scarcity pricing in the Swedish/Nordic balancing market, and
ii) to optionally develop a model for a simulation of scarcity pricing in the Swedish system.

Svenska kraftnät foresees that the scarcity pricing mechanism can be introduced in the form of price adder(s) on balancing capacity and/or the real-time energy price.

The proposal should be consistent with European electricity market regulations and the European electricity market design. Given the differences between European and US market design, the proposal developed in this report does not directly implement a US-style model, but rather proposes an EU-centric approach with adaptations.

The proposed design focuses on measures for integrating scarcity prices in the real-time market design, but it also accounts for the backward propagation of real-time prices to earlier time frames (including measures to improve that backward propagation).

This phase of the project also puts in place a stylized model of the Swedish system which can serve as a basis for simulating the effect of scarcity pricing using real data (as has been the case with the studies of scarcity pricing in Belgium [5, 6, 9]).

In response to this call, N-SIDE has proposed a comprehensive study for the possible application of scarcity pricing in the context of Sweden and more broadly the Nordic system. The following important market design elements have been explored throughout phase 1, and developed in detail in the following chapters.

(i) Shape of ORDC / adder formulas. European markets already have ORDCs. For instance, fixed reserve requirements that are applied in most, if not all, Member States are a specific form of ORDCs: they are inelastic, and they are typically day-ahead demand curves (with no real-time counterparts). Hogan [26] provides a quantitative argument for linking ORDCs to the value of lost load, and the loss of load probability as a function of the remaining available balancing capacity in the system in real time. The basic formula of Hogan can thus be used to compute scarcity “adders”. Nevertheless, there are specific nuances to the computation of these adders, which are currently being investigated in the context of academic research [9], which include the specific choice of VOLL, and whether the remaining balancing capacity on which the LOLP operator is applied refers to pre- or post-activation balancing capacity. Professor Papavasiliou has further collaborated in the past with the CREG and ELIA towards analyzing historical data in Belgium in order to understand the effect of such calibration choices on Belgian scarcity prices [43]. The effect of calibration on adders has also been investigated empirically in the ERCOT market in [44]. This calibration depends on system-specific imbalance distributions, and this topic is therefore analyzed for the Swedish system on the basis of data made available by Svk. These matters are developed in chapter 2 of the present report.

(ii) Where to apply the adders. Scarcity pricing based on ORDC essentially implements co-optimization in real time. This implies that the theoretically best implementation of the mechanism requires scarcity adders to uplift balancing prices and imbalance settlements, and it also requires to use these adders for implementing a real-time market for balancing capacity [12], i.e. a mechanism for settling balancing capacity imbalances. The latter does not exist in the European balancing market, which is a curious peculiarity in the original design of the EU balancing market [12]. The absence of such a mechanism creates a confusion about where the adders should apply. Quantitative arguments [17] have been developed for understanding which designs are capable of properly back-propagating scarcity adders to forward markets, and why suboptimal design choices may induce flexible resources to be pulled out of the balancing market, which would strip the TSO of much-needed flexibility. This topic is investigated in chapter 5 of the present report.

(iii) Affected products. There has been extensive analysis on the specific products on which the adder should apply [5, 13, 9]. Due to its manual activation and delivery duration, mFRR is the better aligned candidate for implementation, since the dispatch instructions are not governed by automatic control and can therefore be aligned in an economically consistent way to the corresponding balancing prices and adders. This topic is investigated in chapter 4 of the present report.

(iv) Cross-border effects. The unilateral implementation of scarcity pricing by a single or a subset of bidding zone(s) when neighboring zones are not following suit is a topic that warrants closer investigation, since it would be desirable to ensure that the system operator is not exposed to financial imbalances as a result of implementing the mechanism. This topic is investigated in [45], and in further detail in the context of the collaboration of Professor Papavasiliou with the Belgian regulatory authority in supporting the Belgian scarcity pricing market design proposal [6]. This topic is examined in chapter 7 of the present report.

(v) Legal and institutional compatibility. The introduction of scarcity adders into balancing prices, imbalance settlement, and the settlement of balancing capacity imbalances (i.e. the implementation of a real-time market for balancing capacity) requires a careful analysis of the extent to which this is permissible by EBGL and other European legal boundary conditions. A basis for the implementation of scarcity pricing which relies on articles 18.4d and 44.3 of the EBGL is developed in [12], and the issue of legal compatibility has been analyzed more deeply in the context of the Belgian market design proposal [6].

It is further important to note that the interaction of the mechanism with pan-European balancing platforms is a key dimension of the analysis. Since scarcity adders are expected to interact with the settlement prices produced by MARI, it is clear that these adders cannot rely on co-optimization, as MARI is an energy-only platform. There is no co-optimization of energy and reserves foreseen in MARI, nevertheless the Texas experience demonstrates that co-optimization is not a necessary condition for the implementation of scarcity pricing. This topic is developed in detail in chapter 6 of the present report.

(vi) Compatibility with CRMs. Scarcity pricing can coexist with CRMs [2, 23]. Although the presence of the mechanism may reduce the scope and amount of missing money that needs to be recovered from CRMs, it is perfectly possible to implement scarcity pricing based on ORDC in tandem with CRMs that are either based on “usual” capacity auctions or that are combined with reliability options. To the extent that this concern is relevant for Svk, this aspect of the design can be analyzed further during the assignment. This topic is developed in detail in chapter 6 of the present report.

1. Principles of Scarcity Pricing and Literature Review

In this section we commence by covering certain basic principles of scarcity pricing. We then proceed to provide an overview of the implementation of scarcity pricing in various European and US markets, thus covering the state of the art on the topic.

1.1 Principles of Scarcity Pricing [23]

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. 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 [24] can be used to establish that this process can result 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 aimed at controlling for the exercise of market power further obviate price formation, and result in so-called missing money [25], [26].

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 [1], 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. 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 our needs for fast-moving “flexible” resources (since renewable resources typically fluctuate beyond human control). This flexibility can originate from various resources: demand response, storage, or fast-moving thermal units (such as combined cycle gas turbines). Scarcity pricing implemented through the introduction of price elasticity in the procurement of reserve capacity exhibits more frequent and less pronounced price spikes than through scarcity pricing based on the trading of energy alone. The resulting “well-behaved” energy price creates a more favorable economic environment for investment.

In order to illustrate these principles, let us consider a simple illustrative example [27]. Consider concretely a system with imbalances distributed according to a normal distribution with mean of 0 MW and a standard deviation of 10 MW. Suppose that two flexible resources bid into the market: BSPA can provide 10 MWh of balancing energy at 20 €/MWh, and BSPB can provide 10 MWh of balancing energy at 50 €/MWh. Suppose, furthermore, that the system experiences an imbalance of 15 MW, which can be interpreted as a price-taking demand for balancing energy.

We pose the following hypothetical question: At this level of stress, what is the value of additional capacity to the system? One way to argue (which is cast on a disciplined quantitative basis in section 1.2) is to notice that, with 5 MW of reserve left in the system, the probability of losing load is $P[Imbalance \geq 5\ MW] = 30.9\%$. One can then infer that, at this level, reserves prevent load shedding 30.9% of the time. Therefore, assuming that VOLL is equal to 1000 €/MWh, one could argue that the value of an additional MW of reserve translates to 1000 €/MWh x 30.9% = 309 €/MWh. However, note that the balancing energy price in this example amounts to 50 €/MWh.

How can one resolve this apparent distance between the value of standby capacity, and the price signal sent by the market? The idea proposed by Hogan [1] is to take the point of view that the real-time market trades not only energy but also balancing capacity, i.e. that it is a *multi-product auction*. Note that future integrated European balancing platforms (MARI and PICASSO) are already multi-product auctions, because they trade both energy but also transmission rights (the latter being traded implicitly). The idea of scarcity pricing is to add reserve into the set of products of our real-time multi-product auctions, in the same way that article 40 of the EBGL [28] considers the introduction of reserve as an explicit product in the European day-ahead multi-product auctions. One then arrives to a co-optimization formulation of the balancing market, as described in detail in section 3.4. We summarize, here, a simplified version of this co-optimization model, where there is no network, in order to highlight the key concepts. The notation used here is identical to that of section 3.4, and is summarized in appendix B. The co-optimization model can be formulated as follows:

$$max_{x\geq 0, xR\geq 0}\ Q_i \cdot P_i \cdot x_i + QR_i \cdot PR_i \cdot xR_i \tag{1.1.1}$$

$$x_i \leq 1 \tag{1.1.2}$$

$$x_i + \frac{QR_i}{Q_i} xR_i \leq 1 \tag{1.1.3}$$

$$(\lambda): \sum_i Q_i \cdot x_i = 0 \tag{1.1.4}$$

$$(\lambda R): \sum_i QR_i \cdot xR_i = 0 \tag{1.1.5}$$

The first term of equation (1.1.1) corresponds to economic welfare generated by the trade of energy, while the second term corresponds to economic welfare generated by the trade of reserve. Constraint (1.1.2) describes the quantity limit on the activation of a BSP offer for balancing energy. Constraint (1.1.3) is a key constraint in the model, and corresponds to the so-called “linking of bids” [11, 29]. Bid linking refers to the fact that energy and reserve bids are dependent on each other due to the fact that a given flexible resource has to allocate its finite generation capacity between the production of energy and idle standby reserve capacity that can be activated on demand. Constraint (1.1.4) is the market clearing condition which requires that the activated balancing energy is equal to the system imbalance. Constraint (1.1.5) is a corresponding market clearing condition for the balancing capacity market, and requires that the supply of balancing capacity by BSPs equal the demand for balancing capacity by the TSO.

In the context of this co-optimization model, we can revisit our simple illustrative example:

- Balancing service provider energy bids (Q_i MWh @ P_i €/MWh) are as follows:
  - BSPA: 10 MWh @ 20 €/MWh
  - BSPB: 10 MWh @ 50 €/MWh
- TSO energy bids:
  - Price-taking for 15 MWh
- Balancing service provider balancing capacity bids (QR_i MWh @ PR_i €/MWh):
  - BSPA: 10 MW @ 0 €/MWh
  - BSPB: 10 MW @ 0 €/MWh
- Based on the implied value of reliability that we discussed previously, we can argue for a TSO balancing capacity demand of 10 MW @ 309 €/MWh. This valuation of reliability reflects the uncertainty that is ever present in operations that take place in advance of real time, even as we near the balancing timeframe.

Applying this data to the market model (1.1) results in a balancing energy price λ of 359 €/MWh, and a balancing capacity price λR of 309 €/MWh. The balancing capacity price is driven by the fact that the TSO demand for balancing capacity is partially satisfied, and thus at the money, which implies that the price of reserve should be set at the valuation of the TSO for reserve. The balancing energy price is implied by the fact that, since BSPB splits its total capacity between the provision of energy and the provision of reserve, this BSP has to be indifferent between its profit margin in the energy and in the reserve market. Thus, the energy price is the marginal cost of this BSP uplifted by the price of reserve, which we refer to later as the scarcity adder.

Note the **pay-for-performance attribute** of the mechanism: flexible assets are rewarded for being present when the system is tight (either by providing balancing energy, or standby capacity). This desirable pay-for-performance attribute of the mechanism is also underlined by the European Parliament in Regulation 2019/943 [22], as we discuss in section 5.1.

Given that co-optimization is not actually implemented or foreseen in the immediate future in European balancing platforms such as MARI, the question becomes how one can approximate the co-optimization outcome. In order to integrate seamlessly (to the extent possible) our proposal with the EU design, while maintaining the intended benefits of scarcity pricing based on operating reserve demand curves (ORDCs), the following three-step procedure has been proposed by Papavasiliou [5, 12], and is inspired by the ERCOT design [4]:

- Run the energy-only balancing platform, i.e. MARI

- Compute the scarcity adder based on[1] how much reserve is available in the system in real time
- Adjust settlement by using the balancing price and the computed adder. How settlement should be adjusted is the focus of section 4.

1.2 Literature Review

European markets already have ORDCs. For instance, fixed reserve requirements that are applied in most, if not all, Member States are a specific form of ORDCs: they are inelastic, and they are typically day-ahead demand curves (with no real-time counterparts). Figure 2 presents an example of an ORDC for mFRR which corresponds to a fixed reserve requirement of 700 MW, as well as a stepped ORDC curve, where we have two different penalty levels as the system becomes increasingly tight.

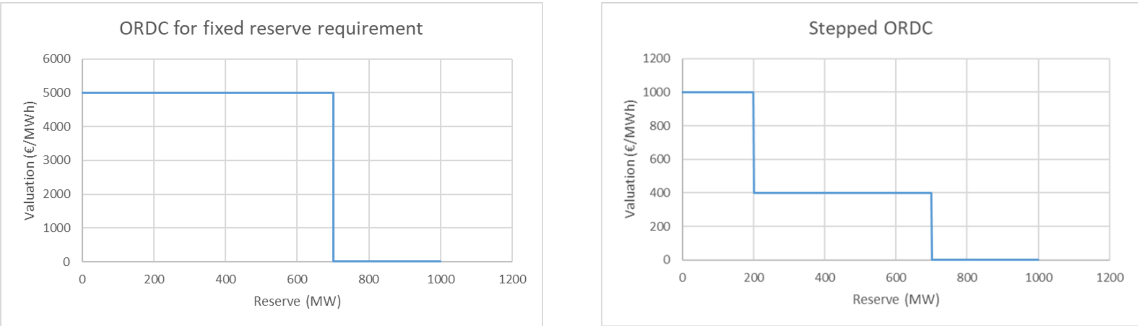

Figure 2: ORDC for fixed reserve requirement of 700 MW (left) and stepped ORDC curve (right).

Hogan [1] provides a quantitative argument for linking ORDCs to the value of lost load, and the loss of load probability as a function of the remaining available balancing capacity in the system in real time. The basic formula of Hogan can thus be used to compute scarcity "adders". The model is based on a two-stage consideration of real-time operations which aims at quantifying the value of reserve in real time. Graphically, the sequence of events of this model is represented in Figure 3.

[1] For this purpose, one can use formula (1.2.7) that is based on LOLP and VOLL considerations and which is justified in section 1.2. Although recommended (and increasingly adopted in various markets throughout Europe and the US), the application of this formula is not necessary.

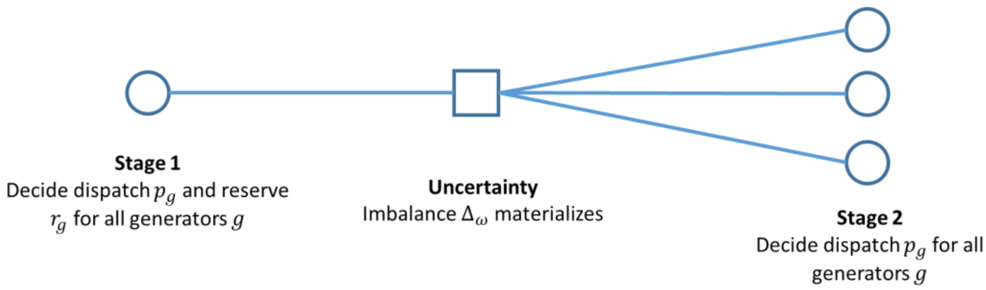

Figure 3: Two-stage model developed by Hogan in order to argue about the shape of an ORDC based on VOLL and LOLP.

The above two-stage stochastic dispatch can be represented in modeling form as follows:

$$\begin{aligned} max_{d,p,r,\delta\geq 0} \int_{x=0}^{d} MB(x)dx - \sum_{g\in G} \int_{x=0}^{p_g} MC_g(x)dx \\ + \sum_{\omega\in\Omega} \Pi_\omega \cdot (VOLL \cdot \delta_\omega - \widehat{MC}(\sum_g p_g) \cdot \delta_\omega) \end{aligned} \tag{1.2.1}$$

$$(\lambda): \sum_{g\in G} p_g \geq d \tag{1.2.2}$$

$$(\mu_\omega): \sum_{g\in G} r_g \geq \delta_\omega, \omega \in \Omega \tag{1.2.3}$$

$$(\rho_g): p_g + r_g \leq P_g, g \in G \tag{1.2.4}$$

$$(\gamma_\omega): \delta_\omega \leq \Delta_\omega, \omega \in \Omega \tag{1.2.5}$$

The objective function of equation (1.2.1) aims at maximizing economic welfare, which is defined as the difference between consumer benefit and production cost resulting from pre-contingency scheduling decisions (first line of the objective function) plus the *expected*[2] consumer benefit minus balancing activation cost after a contingency occurs (second line of the objective function). Constraint (1.2.2) corresponds to energy balance, i.e. it requires that production be equal to demand. Given a realization of a scenario ω, which corresponds to a specific realization of imbalance Δ_ω, constraint (1.2.3) indicates that the amount of imbalance demand that can be covered cannot exceed the amount of available reserve capacity (but it is not guaranteed to be equal to it either, since if the probability of a given realization is too low, it may not be worth scheduling the system so as to serve this level of demand). Constraint (1.2.4) implies that the setpoint plus the amount of reserve capacity that is provided by a unit to the system cannot exceed the nominal capacity of a unit. Finally, constraint (1.2.5) implies that the amount of imbalance energy satisfied in scenario ω cannot exceed the total

[2] Note that the second line of the objective function is an expectation over imbalance realizations, where is the probability of scenario materializing.

imbalance Δ_ω that is realized under this scenario. The notation is presented in detail in the appendix.

We arrive to an ORDC based on VOLL and LOLP using the analytical development of Hogan [1], which is developed in further detail in section 2.1 of [3]. Concretely, let us consider the optimality conditions of a marginal unit g_m, i.e. a unit which is both producing energy as well as offering reserve capacity, i.e. $p_{g_m} > 0$ and $r_{g_m} > 0$. For such a unit, the first-order conditions imply the following:

$$\begin{aligned} \lambda &= MC_{g_m}(p_{g_m}) + \sum_{\omega\in\Omega} \Pi_\omega \cdot \widehat{MC}'(\sum_{g\in G} p_g) \cdot \delta_\omega + \rho_{g_m} \\ &= MC_{g_m}(p_{g_m}) + \sum_{\omega\in\Omega} \Pi_\omega \cdot \widehat{MC}'(\sum_{g\in G} p_g) \cdot \delta_\omega + \sum_{\omega\in\Omega} \mu_\omega \end{aligned} \tag{1.2.6}$$

Using analytical arguments based on the KKT conditions of the problem, we can show that

$$\sum_{\omega\in\Omega} \mu_\omega = LOLP(\sum_{g\in G} r_g) \cdot (VOLL - \widehat{MC}(\sum_{g\in G} p_g)) \tag{1.2.7}$$

The second term of equation (1.2.6) can be ignored as being negligible[3] [1]. And the first term is the usual notion of balancing price, namely the marginal cost of the marginal unit. The analysis concludes that the equilibrium price of energy in a co-optimization is (approximately) equal to the first term plus the third term, i.e. the balancing price plus a scarcity adder which depends on the loss of load probability and VOLL.

Stepped ORDCs correspond to an evolution of fixed reserve requirements. ORDCs based on VOLL and LOLP correspond themselves to evolutions of stepped ORDCs which are based on the theoretical derivation of equation (1.2.7) above. The status of international markets in this evolution is summarized in Table 2, which is sourced from [2].

Note that the entry which concerns Belgium, and in particular the affirmative in the proposal of a real-time market for reserve capacity, is the position of Papavasiliou [5] and the CREG [6], but not ELIA [21]. This dilemma is discussed in further detail in section 4 of the present report. Belgium currently does apply an adder on imbalance prices, the so-called alpha, although this is not a scarcity adder, and is debated within Belgium between ELIA and the CREG, with an intention of the CREG to suppress it, see paragraphs 35-41 of [41].

The Netherlands is considering the introduction of a scarcity pricing mechanism which would apply to both balancing prices and imbalance settlement [42]. Concretely, in paragraph 3.1 of [42] it reads:

[3] Consider, for instance, a system with a linear supply function. Then the derivative of marginal cost is the derivative of a constant, i.e. zero.

“I*n order to support the energy-only market, TenneT NL is considering to include an additional scarcity component to further increase both the common balancing energy price and the imbalance price.*”

It is as of now unclear whether the Dutch scarcity adder would rely on VOLL and LOLP or some other quantities.

Regarding the Polish implementation, the price adder (called COR in Polish) would be computed ex-post and is the payment of non-contracted but available balancing capacity (and only applies to mFRR and RR). The adder is based on the VOLL and the LOLP. It will not be coupled with the balancing price. The LOLP is based on the balancing capacity error forecast.

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 [8, 9]	LOLP/VOLL	No	Yes	The mechanism is proposed, not implemented
Netherlands [60]	Unclear for now	No	No	
Poland [17]	LOLP/VOLL	No	Partially	Planned for implementation by the first half of 2023 Payment of adder to excess real-time mFRR and RR capacity
UK [19]	LOLP/VOLL	No	No	Scarcity adder applied on the imbalance price
Ireland [20, 21]	Steps	No	No	
Greece [16]	LOLP/VOLL	No	Yes	The mechanism is proposed, not implemented
ERCOT [10]	LOLP/VOLL	No	Yes	ERCOT is moving forward with introduction of real-time co-optimization [25]
PJM [10]	LOLP/VOLL	Yes [18]	Yes	Real-time co-optimization for all reserve products to be introduced in 2022 [25]
ISO-NE [10]	Steps	Yes [22, 23]	Yes	Market monitor recently recommended move to ORDC based on LOLP/VOLL [10]
MISO [10]	Steps	Yes [24]	Yes	Market monitor recently recommended move to ORDC based on LOLP/VOLL [10]
CAISO [10]	Steps	Yes [25]	Yes	
SPP [10]	Steps	Yes [26]	Yes	

Table 2: Status of various international markets in terms of ORDC scarcity pricing design choices. Source: [2].

1.3 Specificities of Nordic system operation

In this section we summarize certain specificities of Nordic system operations. These clarifications have been provided to us by Nordic TSOs throughout the workshops that have been held in phase 1. To the extent possible, we connect these remarks to the content that follows.

An important point that has been raised during the workshops is that the Nordic balancing model is targeted towards following uniform rules. This suggests that any market design choices that would be selected in relation to balancing prices, imbalance settlement, and scarcity pricing, would strive to be as aligned as possible. Thus, the larger the area where inconsistencies in cross-border interaction do not occur, the better this can turn out to be for the implementation of market design proposals.

Another interesting point that has been made during the workshops regards whether multi-area reserve dimensioning would account endogenously for sharing and exchange of reserve. It has been clarified that LFC block dimensioning would be determined first, and from there TSOs would optimize by sharing and exchanging.

One additional point to underline is that there is currently a distinction between strategic reserve and disturbance reserve. We provide some information regarding each of these products below, and how they are treated in different Nordic countries.

Disturbance reserve

Finland

Fingrid has clarified that disturbance reserve in Finland (power plants owned or leased by Fingrid that meet the mFRR obligations[4]) are used after the commercial bids, if needed. These reserves are activated at the same price as the last accepted bid in the mFRR market and therefore these do not currently affect balancing / imbalance prices (as long as they are bid at a lower price than the last accepted bid in the balancing market), nor do they highlight scarcity through an increased price.

Thus, the Finnish disturbance reserve is used after commercial bids and can in theory set the mFRR price and thereby imbalance price, but is most often at a lower bid price than the commercial bids, so in practice this does not happen. Note that, in Finland, this reserve is considered part of mFRR.

Sweden

Disturbance reserve in Sweden mainly corresponds to gas turbines that are owned by a company that is affiliated to Svk, or are secured by long-term contracts. In addition, yearly supplementary procurement is performed in order to secure dimensioning. In Sweden, this reserve is used in order to cover for large incidents. It can only be activated when all market bids have been used, and it does not affect any price (mFRR or imbalance price).

There is an ongoing effort in Sweden to transform disturbance reserves, so that they can be in a position to deliver mFRR services. This would then rationalize including them in a scarcity pricing mechanism based on ORDCs for mFRR.

Strategic reserve

Strategic reserves (also referred to as **Effektreserven**, translated as **peak load reserve** in English) cannot receive remuneration from the balancing market or other wholesale markets. According to article 24 of the CEP [22], strategic reserves are only used as a last resort, out of the balancing market, used for congestion, but also if balancing resources are exhausted. There is now a rule with the Recast of the market regulation that requires the following (article 22 of the CEP [22]):

"*during imbalance settlement periods where resources in the strategic reserve are dispatched, imbalances in the market are to be settled at least at the value of lost load or at a higher value than the intraday technical price limit as referred in Article 10(1), whichever is higher;*"

[4] See https://www.fingrid.fi/en/electricity-market/reserves_and_balancing/reserve-power-plants/.

Finland

Strategic reserves would be priced differently (likely with higher prices), but these reserves will cease to exist in Finland for the coming winter, so they will not be relevant in the Finnish case. If activated, this reserve sets the imbalance price to 10000 €/MWh, but does not affect the mFRR price.

Sweden

In Sweden, from the coming winter, strategic reserves will not be activated on the day-ahead market (as they used to be) but as the last resource to exhaust balancing resources according to regulation 2019/943 [22]. When the strategic reserve is activated, the imbalance price will be 1 € above the technical price limit on intraday, so at 10000 €/MWh.

Relation to scarcity pricing

Strategic reserve. The interaction between strategic reserve and scarcity pricing can be captured in the demand side of the scarcity pricing design, in the sense that it can affect the loss of load probability of the operating reserve demand curve. Such a sensitivity has been performed in the context of the discussions regarding the implementation of scarcity pricing in Belgium by Papavasiliou et al. [40], on behalf of the Belgian regulatory authority for energy. Accounting for strategic reserve as a resource that can contribute to the effective capacity carried by the system effectively decreases the loss of load probability in the system, and thus results in an operating reserve demand curve which is shifted to the left and thus produces lower scarcity adders. To put it more directly: when computing the adder in equation (2.1), the Belgian regulator has considered both the possibility of having strategic reserve contributing or not contributing to R in the formula. If strategic reserve is assumed to contribute, then it decreases the adder. On the other hand, this does not mean that strategic reserve needs to be paid the adder, more so if EU legislation prohibits this. The investigation in [40] focused on how the presence of this capacity affected the adder, without assuming that strategic reserve itself was actually paid the adder.

Disturbance reserve. The design that is analyzed in section 4.4 considers a unique price for real-time energy. A blanket application of this principle would imply that resources which qualify for mFRR should be remunerated at a single price, which is also the index for imbalance settlement. If disturbance reserve capacity can be made available in the full activation time of mFRR, then its capacity can contribute to the amount of reserve, R, that is used in equation (2.1). In particular, including disturbance reserve capacity in the formula would increase R, and thus depress scarcity adders, since the system is less tight in real time than it would be in the absence of disturbance reserve, as long as disturbance reserve can respond within the mFRR full activation timeframe.

Svk has explained that disturbance reserve is activated with direct instruction and paid at estimated cost, thus its settlement resembles strategic reserve. This can rationalize accounting for these resources in the calculation of the adder, but not settling them with a scarcity price. It is also worth noting that standard mFRR in Norway does not allow demand response to participate, although EU design is motivating TSOs to convert products into standard products. In fact, an exemption is required for resources to participate as non-standard products.

The above discussion, and the investigation that was pursued in [40], is summarized in Table 3.

	Receives scarcity pricing payments?	Affects adder?	Comments
Strategic reserve	No	Optional [40]	In [40], the sensitivity of the design was tested by adding and removing strategic reserve. The adders decreased when strategic reserve capacity accounted for Sensitivity of this assumption investigated in section 3.5
Disturbance reserve	No (if treated like strategic reserve)	Optional [40]	Sensitivity of this assumption investigated in section 3.6

Table 3: Interaction between strategic reserve and scarcity reserve, as well as disturbance reserve and scarcity pricing.

The design analyzed in section 4.4 is not the only option forward, but it is one that has been analyzed, is well understood, and is based on the economic principles of the law of one price, no arbitrage between energy and balancing capacity, and back-propagation. Alternatives are of course available, but they need to be analyzed individually in terms of economic incentives, and opportunities for arbitrage between different balancing energy products. One thing to always be cautious about in such analyses is to have a genuine distinction between products, if said products are priced differently. Otherwise, agents often find ways to arbitrage, i.e. seek the market that rewards them more handsomely, and dry up the alternative markets that are not keeping up with the equilibrium price of the system. We discuss how this happens in the context of differences between imbalance settlement and mFRR energy in section 4, and analogous analyses could be carried out for alternative designs that may be considered.

2. ORDC Calibration: Specialized Topics

The design that has been proposed for the Belgian market, which is inspired by the ERCOT design [4], relies on updating the imbalance statistics that are used in equation (1.2.7) dynamically. Concretely, the LOLP function depends on the statistics of imbalance for the following reason:

$$LOLP\left(\sum_{g\in G} r_g\right) = \mathbb{P}\left[Imb \geq \sum_{g\in G} r_g\right] = 1 - \Phi_{\mu,\sigma}(\sum_{g\in G} r_g) \tag{2.1}$$

Here, $\Phi_{\mu,\sigma}$ corresponds to the cumulative distribution function of a normal distribution with mean μ and standard deviation σ. Thus, the LOLP depends on the mean and standard deviation of system imbalance (as intuition would suggest). Graphically, the LOLP is represented in Figure 4.

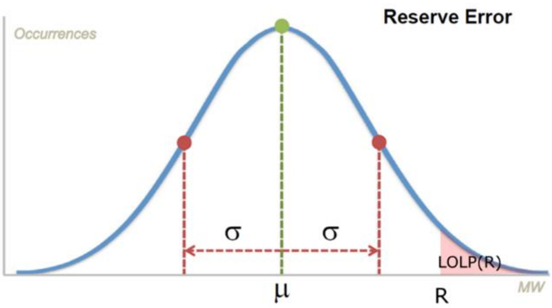

Figure 4: Loss of load probability as a function of the amount of reserve capacity carried by the system. The red area represents the loss of load probability.

A relevant question is how frequently to update the parameters. Inspired by the Texas design [4], for Belgium we propose an updating rule whereby these parameters are estimated for every season of the year, and every 4-hour block of the day, resulting in a total of 24 blocks of imbalance parameter estimates over a given year [5, 6]. The parameters that have been estimated for the Belgian system are presented in Table 4.

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

Table 4: Mean and standard deviation of Belgian imbalances that are used for computing the scarcity adders in the Belgian scarcity pricing design [7].

Another important design dimension is the specific resources that are considered to contribute to reserves of different quality tiers (FRR and RR in particular). Two particularly non-trivial design choices that emerged in the context of the Belgian design were the following:

- Since adders on aFRR were in scope for the Belgian design, it was considered important to decide to what extent different resources could contribute to “fast” reserve of the aFRR type. This proved to be especially crucial in driving the behavior of the mechanism, since the aFRR adder constituted the most important contributor to the overall adder. This is not necessarily a dilemma for the current study, since we focus on a unique adder that is one corresponding to mFRR, i.e. a response time of 15 minutes.
- The contribution of “inter-TSO” reserves had to be decided, and was based on a rather conservative estimate of 50 MW that could reliably be sourced from neighboring TSOs under tight conditions in Belgium.

The overall mapping of resources to reserves of “fast” and “slow” type is indicated in Table 5, and is sourced from [7].

<table>
<tr><th colspan="2">Type of Dispatch and commitment</th><th>Type of plant</th><th>Number of units</th><th>Maximum aggregated production [MW]</th><th>Ability to provide reserve</th></tr>
<tr><td colspan="2">Inelastic</td><td>Cogeneration units
Run of river hydro
Waste incinerators
Nuclear
Classical</td><td></td><td>7 500</td><td>No</td></tr>
<tr><td rowspan="5">Flexible</td><td rowspan="2">Fast balancing capacity</td><td>CCGT</td><td>8</td><td>3 230</td><td>Yes</td></tr>
<tr><td>CCGT-CHP</td><td>2</td><td>524</td><td>Yes</td></tr>
<tr><td rowspan="3">Slow balancing capacity</td><td>OCGT</td><td>6</td><td>302</td><td>Yes</td></tr>
<tr><td>Turbo-jet</td><td>10</td><td>194</td><td>Yes</td></tr>
<tr><td>Non-CIPU</td><td></td><td>250 to 500</td><td>Yes</td></tr>
<tr><td colspan="3">Pump-Hydro</td><td></td><td>1 300</td><td>Yes</td></tr>
<tr><td colspan="3">Inter-TSO</td><td></td><td>50</td><td>Yes</td></tr>
</table>

Table 5: Contribution of different resources in the Belgian system to different types of reserve quality [7].

Even with these design decisions fixed, formula (1.2.7) presented further design questions. These can be summarized as follows:

- What is an appropriate value for VOLL? Higher values of VOLL present a higher ORDC, and a tendency to produce higher scarcity adders.
- In the case of a design with multiple reserve products, what should be assumed regarding imbalance increments[5] within a given imbalance interval? Two extremes can be considered: independent imbalance increments, or perfectly correlated imbalance increments [1]. Since the assumption of independent imbalance increments implies a higher variance for imbalance increments given a total observed 15-minute imbalance, the assumption of independent imbalance increments (which graphically corresponds to a "fatter" ORDC for fast reserves) tends to produce higher scarcity adders for fast reserve. This design dimension is not discussed further in this report, since we commence our analysis with the consideration of a unique mFRR reserve product, but becomes an important design decision.
- Should the leftover reserve in formula (1.2.7) be assumed to be leftover reserve before or after the activation of balancing energy? In systems with finer time resolution (e.g. 5-minute US markets) this distinction becomes less important, but for systems with longer imbalance intervals (e.g. 15-minute EU balancing intervals) this assumption can have important implications. Graphically, assuming that R in equation (1.2.7) corresponds to the post-activation leftover reserve corresponds to a horizontal translation of the ORDC by the imbalance of the interval. Assuming that R is the post-activation leftover reserve tends to increase the adders, since under tight conditions the leftover reserve can become quite small and the LOLP quite high.

[5] If we wish to distinguish fast from slow moving reserves and value them differently, we can split an imbalance interval within further slots of time, e.g. a first and a second 7.5-minute interval. This is the approach, for instance, adopted by Hogan [1], which is later adopted in the ERCOT design [4]. An imbalance increment refers to the delta of the total imbalance that materializes in each of these two steps.

The geometric implication of these assumptions on ORDCs are indicated in Figure 5. The impact of these assumptions on system operation has been analyzed in detail in [9]. This analysis has formed the basis of our recommendation to the Belgian regulatory commission for electricity and gas, whose final decision is published in [6]. This quantitative analysis has been based on a bottom-up unit commitment and economic dispatch model which represent a best-case operation of the system in terms of economic efficiency. The study in [9] exemplifies one of the numerous analytical benefits of an underlying energy-reserve-transmission co-optimization model for investigating various “what-if” scenarios (for instance, “what if we use a more conservative ORDC for implementing scarcity pricing in Sweden?”).

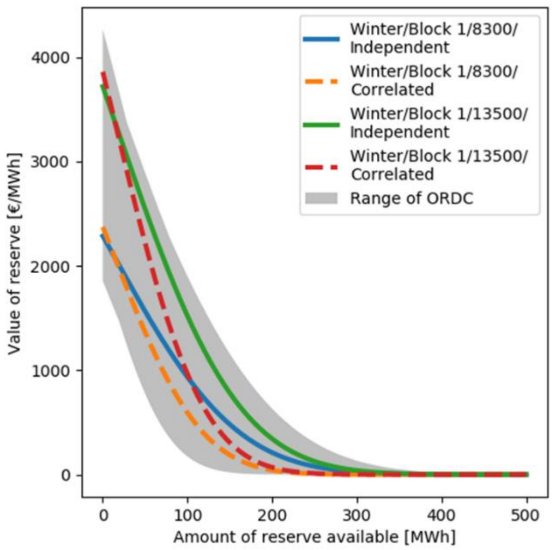

Figure 5: Impact of different assumptions on ORDCs. The grey envelope encompasses all possible ORDCs that can be produced under varying assumptions. Fatter / taller / right-shifted ORDCs tend to produce higher adders, at the expense of more costly system operation [9].

The balance that one seeks to strike with the calibration of the ORDC is one between reliability and cost. Concretely, ORDCs that are taller (high VOLL), wider (independent imbalance increments) or shifted to the right (reserves R are the ones after activation for clearing imbalances) result in more reliable system operation (because the TSO places higher value on a given increment of reserve) but also imply more costly system operation (because the commitment of the required reserves implies that greater min load and startup costs need to be incurred, among others). A system simulator allows us to quantify the tradeoff between these two competing objectives.

Among these three design dilemmas, we note that the second one is not relevant for this study, since we focus on a single type of reserve product, thus rendering the question of imbalance increments irrelevant. On the other hand, the first and third design choice, namely what is an appropriate choice of VOLL and whether or not we should assume to be the pre- or post-activation reserve, are deemed relevant, and appropriate choices can be determined on the basis of quantitative analyses on Swedish data.

These design choices, as well as the extent to which different technologies are assumed to contribute to “fast” reserve, can impact both the cost of operating the system as well as the level of the reserve adder. In terms of the Belgian system, simulation results [9] indicate that the cost of system operation is quite robust to the ORDC design choices and the contribution of different technologies to fast reserve. On the other hand, the level of the scarcity adder does exhibit a noticeable sensitivity to these design choices [9].

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

Table 6: Impact of different design choices and contribution of different technologies to fast reserve on the level of the scarcity adder for the Belgian system. Source: [9].

3. Adaptation of Paradigm to the Swedish / Nordic system

We understand that the scope of the ongoing assignment is to propose an implementation that is applicable to the Nordics. Before tackling this more complex question, we commence by putting forward a proposal for Sweden alone. Our intention is to escalate the analysis to the broader Nordic system in future work.

An important feature of the Swedish system is that mFRR capacity is largely abundant. This can be attributed, to some extent, to the strong presence of hydro resources in Sweden. Although this is true for the overall system, it is also worth commenting that there are specific pockets of the Swedish system that do present a certain degree of reserve shortage, in particular the south of Sweden. This presents interesting and novel challenges related to the implementation of scarcity pricing that have not been encountered in Belgium. Although zonal ORDCs have been analyzed in previous work by Hogan [8], the question has not been investigated in further detail in the European context, to the best of our knowledge, although work is underway in this direction [2]. Before advancing in this more specialized topic, we propose a “vanilla” scarcity pricing methodology for the Swedish system in order to determine the need for a locational instrument.

We commence here with a proof of concept that brings the above concepts to tangible application using Swedish system data. Ultimately, the intention is to expand the analysis to the entire Nordic system. However, we are interested in developing the concept in Sweden only first, for two reasons: (i) we have detailed data available for Sweden from previous interactions with Svk, and (ii) we would like to pursue in further detail the point made by Svk in the kickoff meeting about the fact that Swedish reserves are in relatively short supply.

3.1 Data Input

Svk has provided imbalance data per zone for 2021. The data has been transformed by our team to 15-minute resolution. Note that 32719 entries are available after the transformation of our data, whereas 35040 entries should be available. We fill out the missing values with data that has been provided to use by Svk from previous collaborations[6] which is based on 2018. We use this data to calibrate an ORDC which is used in our subsequent proof of concept. The mean of the imbalances is 28.9 MW, and the standard deviation is 505.4 MW. Applying the standard ORDC formula which is derived in equation 1.2.7 gives us the following:

$$VR(r) = (VOLL - \widehat{MC}) \cdot LOLP(r)$$

We assume a VOLL of 7869 €/MWh, which is based on input from Svk. The graph in Figure 6 shows the ORDC for an assumed marginal cost of marginal unit equal to 0 €/MWh.

[6] Note that the 2021 data that has been provided to us runs up to December 7, 2021. We fill out the missing values with data from January 2018.

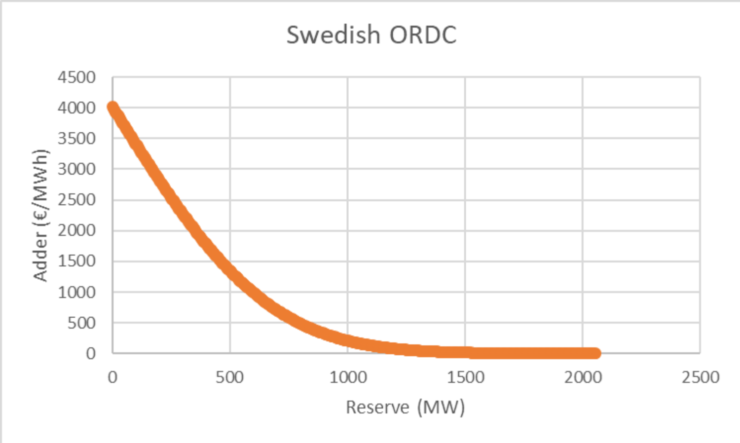

Figure 6: ORDC based on Swedish system imbalances for an assumed system lambda of 0 €/MWh.

Other data input can be summarized as follows:

- Demand data is for 2021, and provided from Svk.
- Generation capacities, variable operating and maintenance costs, emissions rates, and fuel costs are sourced from a third party.
- Efficiencies, ramp rates, and energy storage capacities of reservoirs are sourced from the ten-year network development plant.
- Following information provided by Svk, the energy capacity of each reservoir is as follows: SE1 at 14426.3 GWh, SE2 at 14798.6 GWh, SE3 at 2635.1 TWh, and SE4 at 71.4 GWh. The total capacity of Sweden is[7] thus equal to 31.9 TWh.
- Water inflow time series and wind production time series are sourced from a third party.
- The 2019 price of CO2 is 24 €/tonCO2. We use it as a proxy for 2021.

The capacities and costs of the different Swedish technologies are presented in Table 7. We compute marginal cost as follows from the raw input data:

$$MC_i = \frac{CO2Price[€/tonCO2] \cdot EmmissionsRate_i[tonCO2/MWh]}{Efficiency_i} + \frac{FuelCost_{Fuel(i)}\left[\frac{€}{MWh}\right]}{Efficiency_i} + VOM[€/MWh]$$

We have third-party data for the average availability of different technologies. For hydro, we override this data with the data given by Svk, whereby out of the 16334 MW of hydro capacity only 12200 MW is effectively usable[8], due to various environmental

[7] Based on communication with Svk, it has been explained that the theoretical energy capacity of Sweden is equal to 33.7 TWh, from zero to full storage. However, this differs from the amount of energy that is actually usable, and which is rather closer to 29 TWh. Following discussions with Svk, we use an intermediate value of 31.9 TWh in the energy capacity rating of the reservoirs in our model.

[8] Based on information received by Svk, the nominal rating of hydro resources is actually in the range of 13.2 GW. However, this value is rather an upper bound, as it requires short- and long-term dispatch strategies to be aligned. In a typical year, the rating is instead in the range of 12 GW – 12.4 GW, hence our choice of 12.2 GW. This discounting allows us to indirectly account for aggregation approximation

and other constraints, which implies an availability of 80.8% for hydro generators. VOM in the above equation stands for variable operating and maintenance costs. $EfficiencyRate$ stands for the by which the primary fuel of each technology is converted into electricity.

Based on information received by Svk which is sourced from Nordpool spot, we assume that reservoirs commence from a storage level of 24.8 TWh in the beginning of the year, and end with a storage level of 19.6 TWh.

In order to better align the results of our model to the average prices that were observed in Sweden in 2021, we have included linear supply functions at every zone of the model, which correspond to export price elasticity. This introduces a certain degree of price variability in the model beyond the marginal costs of the technologies that are listed in table 7.

In order to calibrate these "border supply functions", we first conduct a sensitivity analysis which leads to the following interesting observation: ranging the exports from 100% to 70% of their value leads to a drop in average prices from approximately 120 €/MWh to approximately 35 €/MWh, with the drop occurring abruptly at approximately 75% of exports. Since historical average energy prices in 2021 have ranged in at around 60-70 €/MWh, we introduce a linear supply function that starts at 0 €/MWh and increases up to 120 €/MWh at 30% of the average exported energy throughout the year. This export price elasticity allows the prices to depressurize, but not excessively.

Concretely, for the single-area model we have a "border supply function" that rises up to 120 €/MWh at 898 MW. For the multi-area model, it is observed that SE2 actually imports on average, so its "border supply function" is set to 0 MW, whereas the border supply functions of the other zones rise to 180 €/MWh at the following capacities: SE1 at 290 MW, SE3 at 192 MW, and SE4 at 432 MW.

3.2 Single-Area Model

The energy storage level in the single-area model is indicated in Figure 7. The average price in the model is 73.3 €/MWh.

errors, as well as limited water rights in the summer. Limited water rights imply that, even if abundant capacity is available nominally, there are not enough rights to generate hydropower. Hydropower production needs to be planned ahead, and reservoirs should have optimal head levels at all plants, and optimal levels in storage stations. Water rights correspond to permits, which indicate a typical flow that needs to be discharged (e.g. water rights can stipulate that "*from January until September x cubic meters per second should be discharged, and water cannot go above or below certain head levels*". These water rights are related to the need of preserving the physical environment for recreation activities, preserving biodiversity, controlling flooding, and other purposes. Actual hydropower capacity in the summer may even be as low as approximately 10-11 GW. This is anyway difficult to know on the basis of historical data, because in the past hydropower units in Sweden have never produced at their full capacity during the summer, since market prices are lower during these periods. Winter months are more likely to reflect the true capacity of the hydropower plants, if one attempts to infer this capacity from historical data.

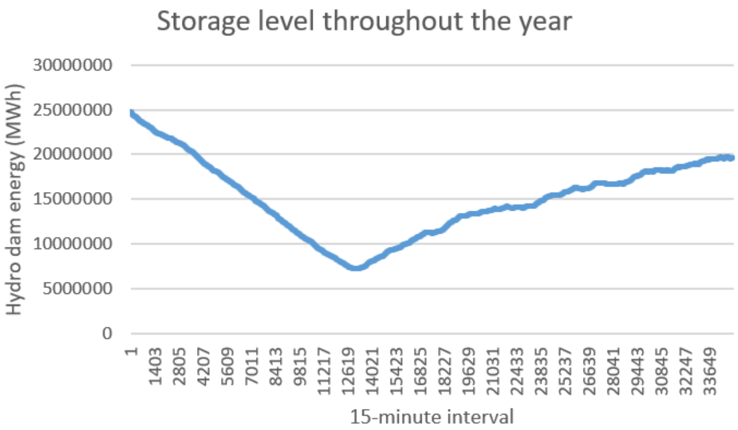

Figure 7: Energy storage level throughout the year 2021.

The average reserve prices are equal to 0.39 €/MWh. There are 560 periods with non-zero scarcity prices. The reserve price duration curve is presented in Figure 8.

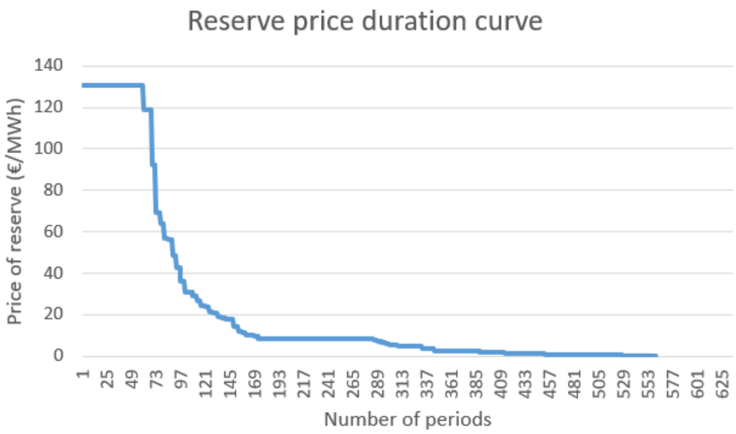

Figure 8: Reserve price duration curve for the single-area model.

The occurrence of non-zero scarcity adders in January and February can be attributed to the peak net demand which occurs during these months, as indicated in Figure 9.

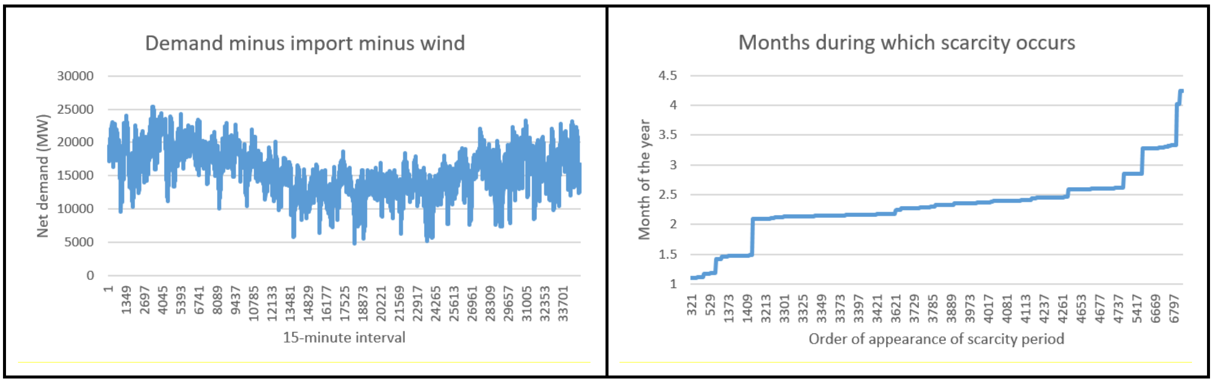

Figure 9: Net demand (left) and month of occurrence of scarcity (right) in the single-area model, where we can observe that January and February correspond to peak-loaded periods for the system.

The link between system-wide scarcity and net demand in the system becomes evident in Figure 10. As we can observe in the Figure 10, reserve adders are largely driven by system net demand, thus confirming that scarcity pricing can also function as an instrument that can contribute towards tackling system adequacy, in addition to flexibility. Average prices during scarcity periods are equal to 120.4 €/MWh.

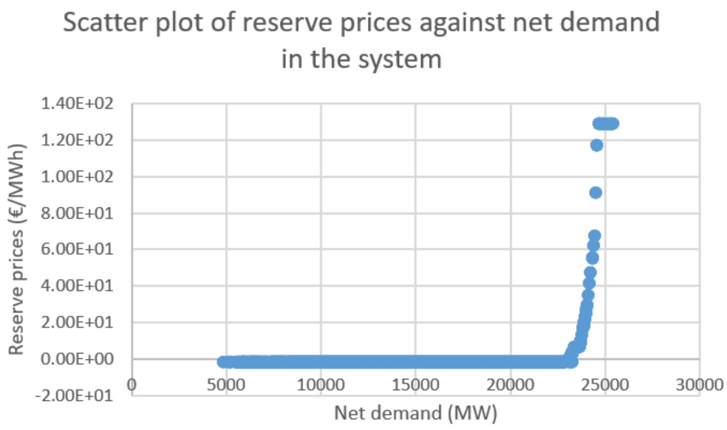

Figure 10: Scatter plot of reserve prices versus net demand in Sweden. The hockey stick increase in the right of the graph corresponds to system-wide shortage in January, February and March.

3.3 Multi-Area Model

We now proceed to a multi-area model. For this, we need to know what the installed capacity is at every zone, along with the demand at every zone and the network data. We proceed as follows:

- Demand data is available per zone for 2021 by Svk
- ATC data is available between bidding zones for 2021 by Svk. Missing values are filled out with the data of the previous time step. From this ATC data, we can compute interconnector capacities.
- It is clearly pointed out by Svk that all nuclear capacity is in SE3.
- The hydro capacities are based on table 3 of [10]. We further use the inflow profile provided to us by a third party multiplied by the ratio of installed hydro capacity in each Swedish zone in order to infer a storage inflow time series per zone. It has been communicated to us by Svk that the storage inflows for 2021 amount to 67.8 TWh, thus the original storage inflow time series has been scaled up to reflect this total inflow over the year.
- We attribute the “OCGT_A” technology of our data to the “Gasturbiner + övrigt” technology of table 3 of [10]
- We ignore the “Kraftvärme, fjärrvärme” and “Kraftvärme, industri” technologies of table 3 of [10], since these technologies are anyway following a heating profile and are therefore not flexible.
- We introduce a new “Condenser” technology to the model, and set its marginal cost to 180 €/MWh, following guidance by Svk. We assign an availability of 89% to this resource, following the OCGT availability (although availability data is

not used in our default proof of concept, since it leads to systematic shortage, as indicated previously). Note that, in Sweden, this technology acts as strategic reserve[9].

- Imports/exports are provided by Svk, and sourced from NordPool. The topology of the interconnections of the Swedish system are indicated in Figure 11. Intraday values are added to the day-ahead values.
- For wind production we have the following link giving us annual TWh production in Sweden: https://www.statista.com/statistics/737840/electricity-from-wind-production-sweden/. We then use profiles of wind production (Sweden-wide) from a third party. We use the data provided by Svk regarding weekly production of wind in each Swedish zone to partition the profile given by a third party per zone.

The spatial distribution of capacities in the multi-area model, along with their marginal cost, is presented in Table 7. We note that, in practice, icing on rivers causes reduced availability of hydro capacity. For instance, 2 GW of hydro capacity was unavailable in SE1 and SE2 due to icing, for an entire weekend in November of 2021. This feature can be added straightforwardly to our model, and would result in higher scarcity adders than what is observed currently.

Technology	Marg cost	Nominal capacity SE1 (MW)	Nominal capacity SE2 (MW)	Nominal capacity SE3 (MW)	Nominal capacity SE4 (MW)	Total nominal capacity (MW)
Condenser	180	0	0	243	662	905
Hydro dam	2.7	5320	8076	2593	345	16334
Nuclear	14.2	0	0	6871	0	6871
OCGT	49.6	1	2	962	618	1583
Wind onshore	0	1652	3876	2891	1598	10017

Table 7: Spatial allocation and marginal cost of different technologies of the Swedish system in our multi-area model.

[9] In principle, strategic reserve is used as an out-of-market intervention and is accompanied by specific pricing protocols in different Member States. We ignore this aspect in the present analysis, and assume that condensers are part of the normal merit order.

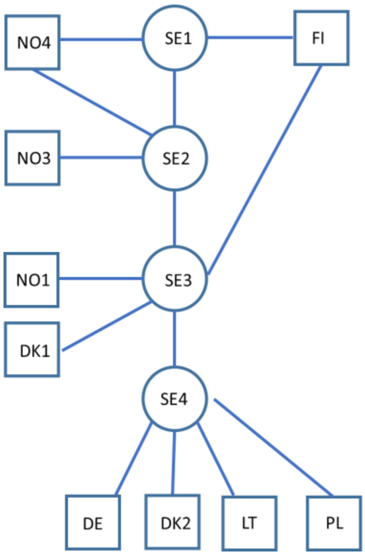

Figure 11: Interconnections of Swedish system to zones out of Sweden. Circles indicate Swedish zones, boxes indicate zones out of Sweden.

3.4 ORDC in a Multi-Area Model

Let us consider a unique reserve requirement in SE4 as the simplest first step in our analysis. We present a co-optimization model for this case, discuss pricing, and then discuss how one could emulate the outcome of the co-optimization through ex post adders.

A co-optimization model of energy and reserves on a network can be expressed as follows:

$$max_{d,p,r,dr,f,fR} VOLL \cdot \sum_{z \in Z} d_z - \sum_{g \in G} MC_g \cdot p_g + \int_{x=0}^{dr_z} ORDC_z(x) dx \quad (3.5.1)$$

$$(\lambda_z): d_z - \sum_{g \in G_z} p_g + \sum_{k=(z,\cdot) \in K} f_k - \sum_{k=(\cdot,z) \in K} f_k = 0 \quad (3.5.2)$$

$$(\lambda R_z): dR_z - \sum_{g \in G_z} r_g + \sum_{k=(z,\cdot) \in K} (fR_k^+ - fR_k^-) - \sum_{k=(\cdot,z) \in K} (fR_k^+ - fR_k^-) = 0, z \in Z \quad (3.5.3)$$

$$(\mu_g): p_g + r_g \leq P_g, g \in G \quad (3.5.4)$$

$$(\gamma_g): r_g \leq R_g, g \in G \quad (3.5.5)$$

$$(\nu_z): d_z \leq D_z, z \in Z \quad (3.5.6)$$

$$(\lambda_k^+): f_k + fR_k^+ \leq T_k, k \in K \quad (3.5.7)$$

$$(\lambda_k^-): -T_k \leq f_k - fR_k^-, k \in K \quad (3.5.8)$$

$$d, p, dr, r, fR^{+/-} \geq 0 \quad (3.5.9)$$

The objective function presented in term (3.5.1) corresponds to economic welfare that consists of economic welfare generated by the trading of energy (the two first terms of equation (3.5.1)) and economic welfare generated by the trading of balancing capacity (the third term of equation (3.5.1)). Constraint (3.5.2) corresponds to the market clearing condition for energy, and is applied per zone. Similarly, constraint (3.5.3) is a market clearing condition for balancing capacity, and it too applies per zone, which implies a separate price for balancing capacity per zone. Constraint (3.5.4) is a "linking of bids" type of constraint, which requires that the amount of energy and balancing capacity traded by a flexible resource not exceed its physical capacity, meaning that the flexible resource cannot double-book its capacity in both the energy and balancing capacity market. Constraint (3.5.5) expresses the fact that inflexible resources cannot provide balancing capacity. Constraint (3.5.6) is the quantity limit on how much energy is consumed. Constraints (3.5.7) and (3.5.8) correspond to the upward and downward limits of available transmission capacity respectively, while constraint (3.5.9) determines the non-negative variables in the problem (and we can note that flows are free variables).

Here, we use the so-called inscribed boxes formulation for the exchange of balancing capacity [11, 29]. The idea is to define a "flow" of balancing capacity in the market clearing model, but in such a way that balancing capacity flows do not net out. We prevent the netting out of balancing capacity flows because there is inherent uncertainty about whether these flows are activated or not in real time. The underlying formulation is based on results from computational geometry [34]. The concept has a precedent in tackling "non-intuitive" flows in the day-ahead energy market, and in the computation of ATC capacities in the intraday market.

Note that the model in this section ignores time indexing. The actual model is in fact dynamic, since hydro storage is involved, but the temporal dimension does not add to the comprehension of the interactions between pricing and ORDC.

We proceed to analyze the KKT conditions of this problem, in order to gain an economic intuition about the behavior of prices, and in order to propose a process for approximating the outcome of co-optimization with ex-post adders. These KKT conditions can be expressed as follows, where we do not repeat the equality constraints for the sake of brevity.

$$0 \leq \mu_g \perp P_g - p_g - r_g \geq 0, g \in G \tag{3.5.10}$$

$$0 \leq \gamma_g \perp R_g - r_g \geq 0, g \in G \tag{3.5.11}$$

$$0 \leq \nu_z \perp D_z - d_z \geq 0, z \in Z \tag{3.5.12}$$

$$0 \leq \lambda_k^+ \perp T_k - f_k - fR_k^+ \geq 0, k \in K \tag{3.5.13}$$

$$0 \leq \lambda_k^- \perp T_k + f_k - fR_k^- \geq 0, k \in K \tag{3.5.14}$$

$$0 \leq p_g \perp MC_g + \mu_g - \lambda_{z(g)} \geq 0, g \in G \tag{3.5.15}$$

$$0 \leq r_g \perp -\lambda R_{z(g)} + \mu_g + \gamma_g \geq 0, g \in G \tag{3.5.16}$$

$$0 \leq d_z \perp -VOLL + v_z \geq 0, z \in Z \tag{3.5.17}$$

$$0 \leq dr_z \perp -ORDC_z(dr_z) + \lambda R_z \geq 0, z \in Z \tag{3.5.18}$$

$$0 \leq fR_k^+ \perp \lambda R_{From(k)} - \lambda R_{To(k)} + \lambda_k^+ \geq 0, k \in K \tag{3.5.19}$$

$$0 \leq fR_k^- \perp -\lambda R_{From(k)} + \lambda R_{To(k)} + \lambda_k^- \geq 0, k \in K \tag{3.5.20}$$

$$(f_k): \lambda_{From(k)} - \lambda_{To(k)} + \lambda_k^+ - \lambda_k^- = 0, k \in K \tag{3.5.21}$$

In order to understand the behavior of scarcity prices under tight conditions, we consider the following scenarios:

- Scenario A: SE4 has reserve scarcity, and none of the links are congested
- Scenario B: SE4 has reserve scarcity, and SE3-SE4 is congested

Under scenario A, we have $\lambda R_z = ORDC_{SE4}(dr_{SE4})$ from (3.5.18), since the reserve demand is not fully covered, thus at the money, and therefore sets the price for reserve at a given zone. Since there is shortage[10], this implies that there is a generator for which $r_g > 0$ and $p_g > 0$, meaning that there will be no generator "sitting around" and thus that at least one generator will be splitting its capacity between energy and reserve. This in turn implies that $\lambda_{z(g)} = \lambda R_{z(g)} + MC_g$ (assuming also that flexible units have non-binding reserve capacity constraints, i.e. $r_g < R_g$), which means that the profit margin in the energy and reserve markets is equal for the zone in question. And since none of the network constraints in the system are binding, we have $\lambda_k^+ = \lambda_k^- = 0$ (i.e. the congestion rents are zero), which implies from (3.5.21) that there is a unique energy price λ in the entire system and also that there is a unique price from reserve since the reserve price is just $\lambda - MC_g$ where g is the marginal unit.

Under scenario B, we have $\lambda R_{SE4} = ORDC_{SE4}(dr_{SE4})$ again from (3.5.18). By the same reasoning as in the previous case, we further have that $\lambda_{SE4} = \lambda R_{SE4} + MC_g$, where g is the marginal unit in SE4. And assuming that the rest of the system is not experiencing shortage, we have that $\lambda R_z = 0$ for the other zones (by definition of what it means to not experience shortage in the rest of the system). Assuming that none of the other links in the SE1-SE3 system are congested, we then have a unique energy price, which is equal to the marginal cost of the marginal unit in SE1-SE3. And this then implies that $\lambda_{SE3-SE4}^+ = \lambda_{SE4} - \lambda_{SE3}$. The rationale for BSPs in the North not being paid despite the fact that they provide balancing capacity to resources in the South is that, by assumption in this scenario, the Northern BSPs are abundant, and the limiting factor is the ATC capacity that can get balancing capacity to the South. Thus, a zero price is compatible with making balancing capacity available from Northern units, since the BSPs that do provide it do not face any opportunity cost.

[10] This follows from the definition of scarcity. There exists a generator in the system, the capacity of which is not in excess. It is cheap enough to produce, but the system is so tight that it is also one of the more expensive units in the system, i.e. one which provides reserve.

The case of involuntary load shedding is a special case of scenario A or B: ORDCs can be designed to have a high valuation, as indicated in Figure 12. The idea of the ORDC that is presented in this figure is that it prioritizes reserve over load service, as long as the valuation (10000 \$/MWh in the hypothetical ORDC of the figure) exceeds the VOLL. In such a case, the co-optimization resorts to involuntary load shedding before it depletes reserve. This is actually applied in Texas: the motivation is that ERCOT prefers to resort to involuntary load shedding instead of putting the system in a highly vulnerable position by depleting reserves of last resort. In such a case, let us consider what would happen in scenario A. Since there is involuntary load shedding and no congestion in the system, the price of energy becomes uniformly equal to VOLL. The price of reserve becomes VOLL plus the marginal cost of the marginal unit, and since it is higher than that of energy, all resources which have a higher marginal cost than the marginal unit are voluntarily willing to keep their capacity aside for the purpose of providing reserve.

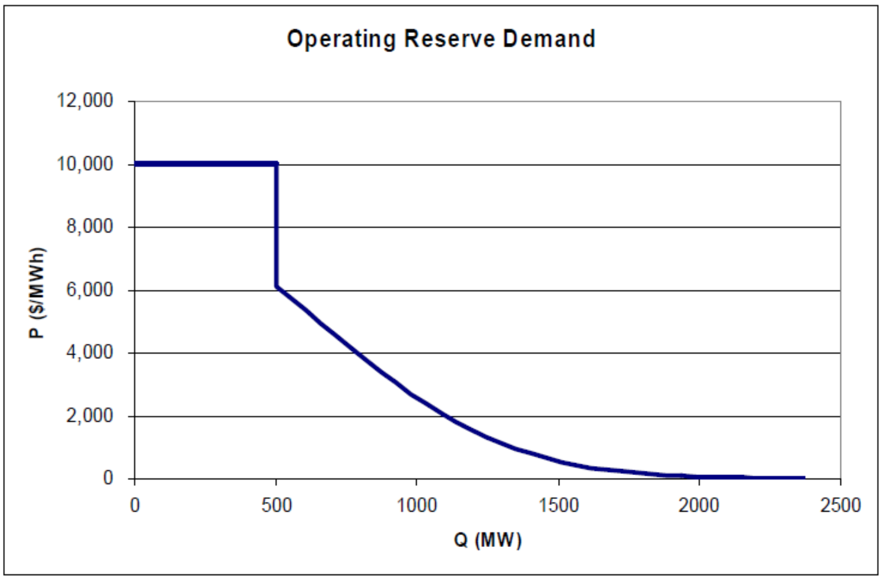

Figure 12: An ORDC with an inelastic component up to 500 MW, which can be used for ensuring a minimum amount of reserves is kept in the system. Source: [1].

We can apply the same logic in order to deal with other cases. For instance, suppose that the congested link is not SE3-SE4, but another upstream link. Then the energy price is equal in the downstream subnetwork, it is also equal in the upstream subnetwork, and the two energy prices are separated by the scarcity adder.

We can thus propose the following approximation to the result of a co-optimization when there is no explicit co-optimization in MARI:

- Compute $\lambda R_{SE4} = ORDC_{SE4}(dr_{SE4})$
- If none of the Swedish network is congested, apply an adder λR_{SE4} to the uniform price of the entire Swedish area
- If the link SE3-SE4 is congested, apply an adder λR_{SE4} to SE4 only
- If another downstream link is congested, apply the adder to the downstream uncongested subnetwork

Each of the scenarios is indicated graphically in Figure 13. We note that, when there is no congestion (left panel), the scarcity adder applies universally to all Swedish zones. On the other hand, when there is congestion (right panel), scarcity adders apply only to the zones which belong to the congested load pocket.

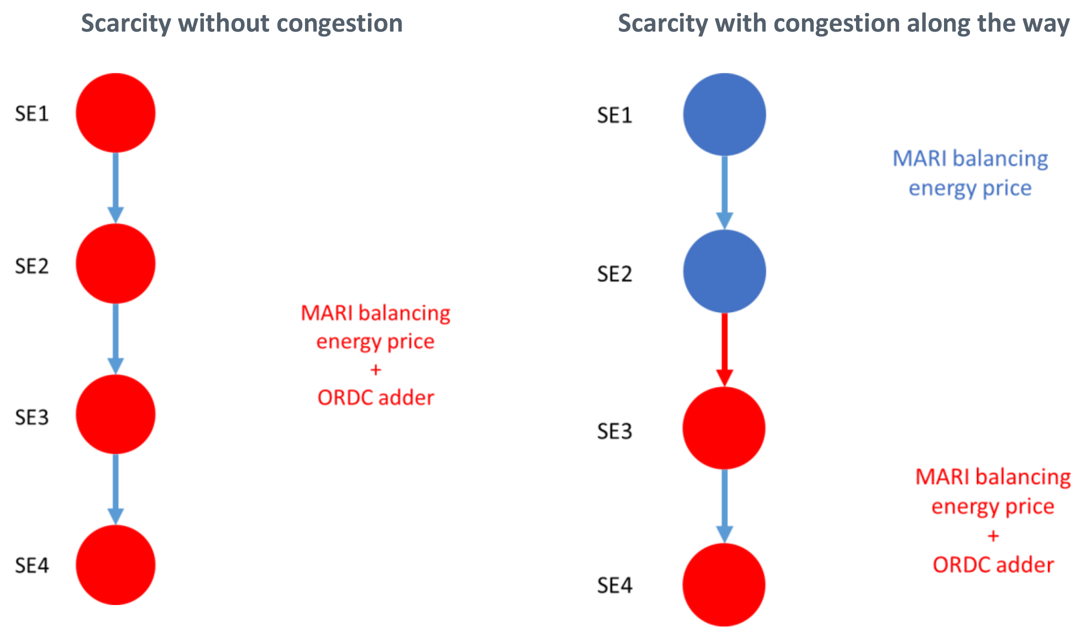

Figure 13: Scarcity pricing in the multi-zone system of Sweden. Red circles represent nodes which experience a scarcity adder, whereas blue circles represent nodes which do not experience a scarcity adder. Blue arrows indicate links which do not experience congestion, whereas red arrows indicate links which do experience congestion.

It is interesting to note that congestion can actually result in a drop of reserve prices. To see why this is the case, let us consider the example that is presented in Figure 14. In the left of the figure we have a situation where the BSP in the North is exporting both energy and reserve to the south. The black fonts correspond to the input data, whereas the blue fonts correspond to the solution of the optimal dispatch. In the case without congestion (left), the BSP in the north produces 1 MWh of energy, and 4 MW-h of reserve, which are exported to the south. The line is not congested, and the price of reserve is 100 €/MWh throughout the entire system. Reserves are scarce system-wide, and this is reflected in the price for reserve in the North. In the case with congestion (right), we decrease the capacity of the interconnector from 6 MW to 4 MW. The line limit becomes binding, and it is now no longer the BSP capacity in the north that is scarce, but rather the capacity of the interconnector. There is a separation of reserve prices. The price of reserve in the North specifically becomes equal to 0 €/MWh, because now there is abundant reserve in the North. The BSP in the North is not binding in terms of its capacity, and thus the price signal that is needed for inducing it to provide a reserve of 1 3 MW-h (at the money in the reserve product) is exactly 0 €/MWh.

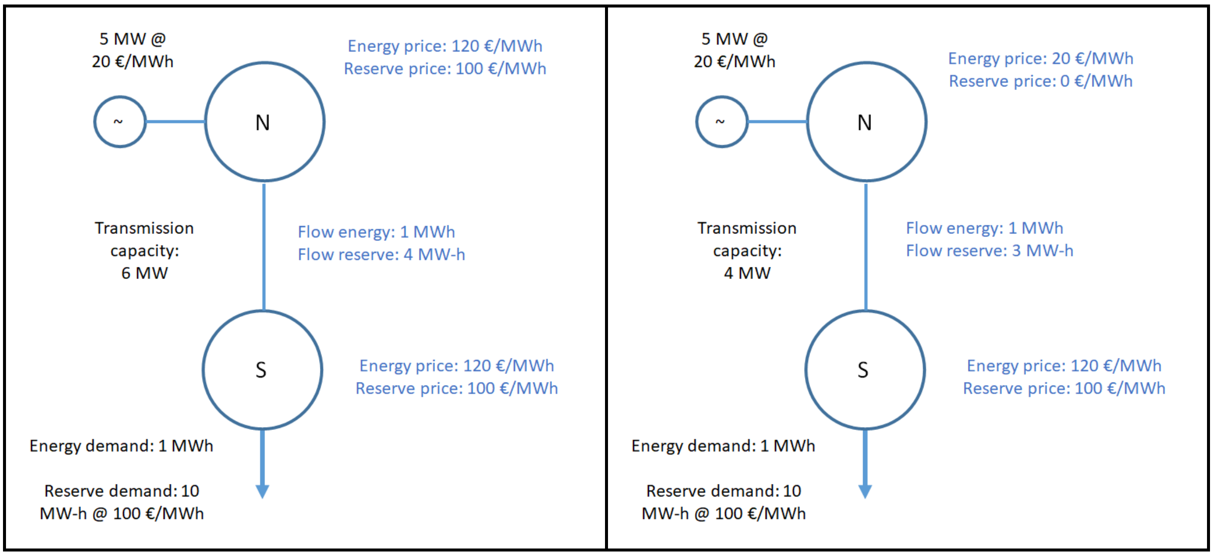

Figure 14: A system in which we transition from no congestion (left) to congestion (right). Note that the price of reserve decreases when congestion occurs, because in this case the north has abundant reserves.

There is no mathematical contradiction in this result. What convex analysis guarantees is that the value function of a convex optimization problem is convex, which implies that if we increase the right hand side of a given constraint, then the dual multiplier of this constraint is increasing. One can use this property to link the market clearing price of a convex model to the amount of demand that appears in the right hand side of a corresponding constraint. Drawing connections between the market clearing price of any constraint of the model (e.g. the price of the reserve market clearing constraint in one zone) and the right-hand side of any other constraint (e.g. the capacity of the link of any given line in the system) is less straightforward.

We proceed next to analyze the results of the co-optimization. We then compare the resulting prices with those produced by our proposed approximation of the co-optimization prices.

3.4.1 ORDC in SE4 Only

In this section we run the full-horizon simulation with the ORDC of Figure 6 applying to SE4 only. We analyze the behavior of stored hydro energy, energy and reserve prices, congestion, and scarcity incidents. Note that, even if there were no congestion, the results of this section would not be identical to those of the single-area model, since the ORDC used in the system-wide model has a different shape from the SE4-only ORDC, which has been calibrated against SE4 imbalances only.

Zone	Initial storage (TWh)	Final storage (TWh)
SE1	11.204	8.855
SE2	11.494	9.084
SE3	2.047	1.617
SE4	0.055	0.044

Table 8: Initial and final storage levels for hydro in the different Swedish zones.

Figure 15 presents the energy stored in the hydro reservoir of each Swedish zone. As mentioned previously, based on information received by Svk which is sourced from Nordpool spot, we assume that reservoirs commence from a storage level of 24.8 TWh in the beginning of the year, and end with a storage level of 19.6 TWh. These initial and final storage levels are split between each zone in proportion to the energy capacity of each reservoir. The resulting initial and final storage levels of each zone that are assumed for the simulation are presented in Table 8. In practice, hydro dams are at their lowest just before the beginning of the spring floods (around week 19-20), depending also on the specific location in Sweden. During the summer, hydro dams cannot absorb too much power. This is due to the fact that there are rules regarding the water levels in the lakes. These rules limit the daily changes in the water levels. For instance, it is not possible to draw power all day long in the summer. Such constraints are not captured in our model[11], but they are anyway not affecting our core analysis. Note that such constraints do not exist during winter time.

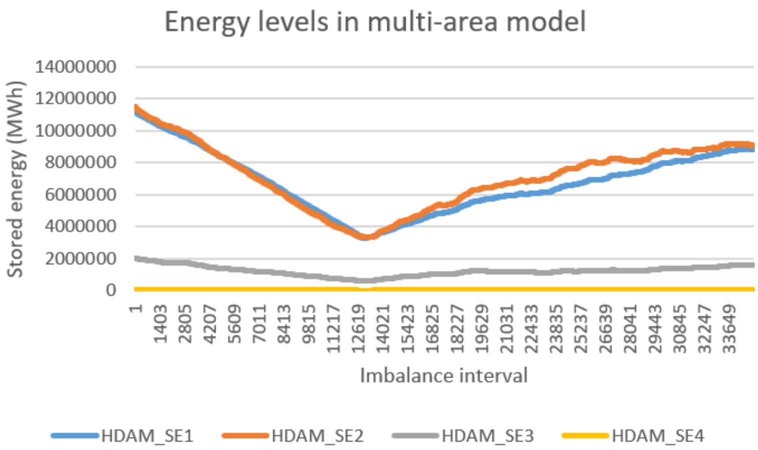

Figure 15: Stored energy in the hydro reservoirs of the multi-area model.

There has been an attempt to align the hydro levels produced by the model with hydro levels that are observed historically, and which are indicated in Figure 16. Concretely, we have introduced a constraint which requires that the hydro storage levels range between 15% and 25% of the energy reservoir storage capacity on April 15th, a time

---

[11] Once water levels are accounted for, there are additional features that are important to model, and which are not captured in our analysis (but which also do not affect the essence of our analysis, which is about short-term balancing, and not about mi-term energy planning). Concretely, due to the shape of the water reservoirs, when hydro dams are close to being full, withdrawing water does not affect the level of water significantly, because the top of the dam covers a large surface. On the other hand, the water level is affected more strongly when the reservoir is depleted, because the dam becomes more narrow the deeper we deplete it.

It is also worth noting that the actual power capacity of hydroelectric generators is equal to 16.2 GW. Nevertheless, when Svk calculates the power capacity of its hydroelectric generators, it typically uses the value of 13.2 GW. This is due to the fact that certain machines may not be running, due to water level constraints, and due to other factors which affect how much storage capacity is available during the year.

around which reservoirs appear to attain their minimum storage levels. This is indeed reflected in the trajectory of the hydro energy levels in Figure 15.

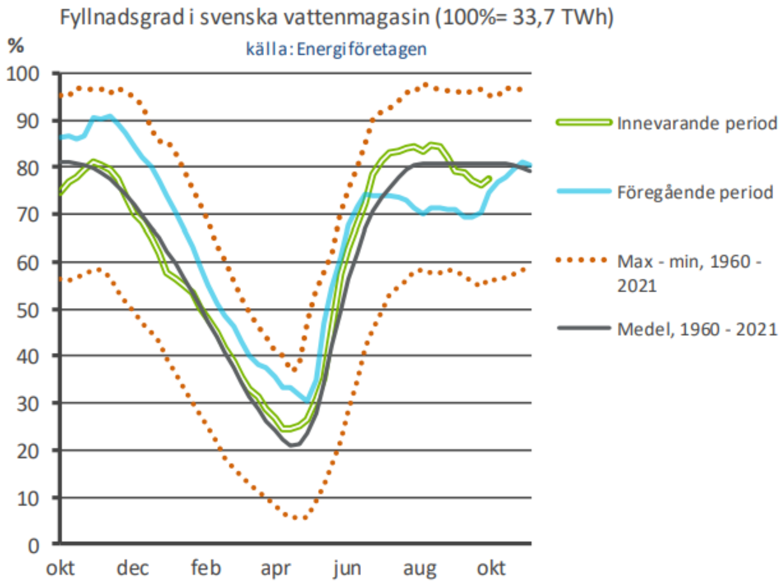

Figure 16: Historically observed levels of hydro energy storage in Swedish reservoirs. Source: https://www.energiforetagen.se/globalassets/energiforetagen/statistik/kraftlaget/aktuellt-magasinslage-sverige-veckorapport.pdf?v=59oGtH9U2Falx-q0iof2xKugd6Y.

We find that the energy price is 71.72 €/MWh in SE1-SE3 and 71.74 €/MWh in SE4. There are no periods during which load shedding occurs.

There are 124 periods during which the scarcity adder is non-zero. The average value of the adder is 0.09 €/MWh in zone SE4 and 0.07 €/MWh in all other zones of Sweden. These scarcity periods correspond to two types of periods (as our theoretical analysis predicts):

- There are periods of scarcity during which the link SE3-SE4 is congested (52 periods). During these periods, we have a clearing price for zone SE4 which is larger than the clearing price of the rest of the system. The average price of reserve in SE4 during this period is 14.43 €/MWh.
- There are periods of scarcity during which the link SE3-SE4 is not congested (72 periods). During these periods, we have a clearing price for zone SE4 which is equal to the price for the rest of the system. The average price of reserve during these periods is, on average, 32.33 €/MWh. This indicates that these periods correspond to tighter conditions for the overall system.

We plot the duration curve of reserve prices in SE4 in Figure 17.

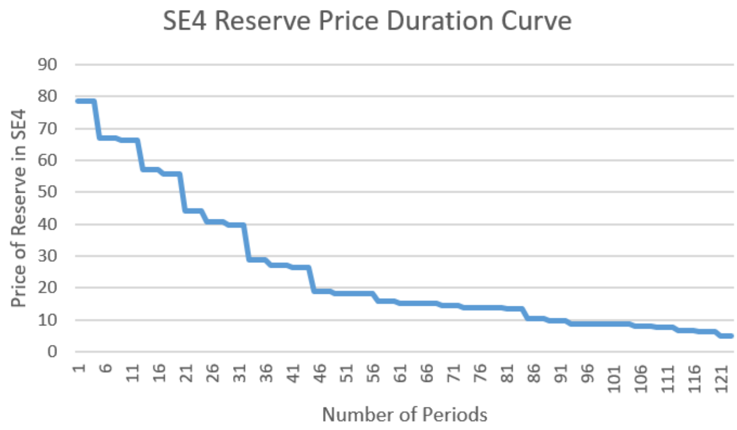

Figure 17: Duration curve of SE4 reserve price in the case where only SE4 applies an ORDC.

In Figure 18 we present the periods during which scarcity is observed in the system. The horizontal axis of the figure is the order of appearance of a scarcity period (for instance, the first period when scarcity was observed in the system corresponds to a value of 1 in the horizontal axis, and so on). The vertical axis is the month during which this scarcity incident appeared. We observe that the predominant months when tightness is observed are February and August, and that there is no scarcity before February or after September. The February months correspond to high net load (load minus imports minus wind), as indicated in Figure 19. During these periods, scarcity occurs system-wide. The August months correspond to congestion in the SE3-SE4 link. During these periods, the system is indeed not as heavily loaded as in February. In Figure 20 we present the scatter plot of net demand in the system against reserve prices in SE4. We indeed observe one cluster of non-zero reserve prices which originate from system-wide scarcity (in February), and another cluster of non-zero prices during periods when the SE3SE4 link is congested (August).

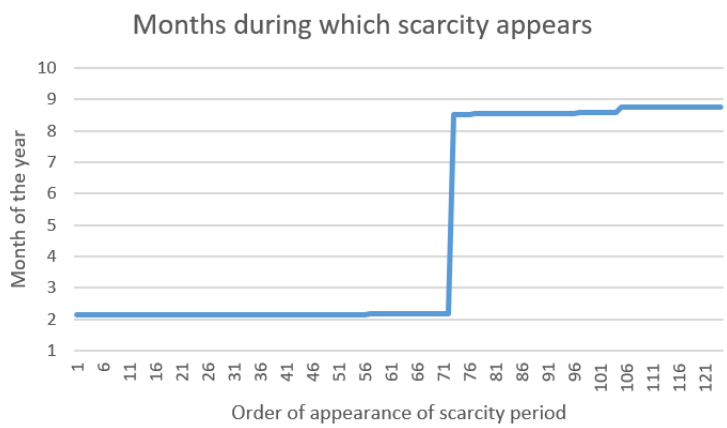

Figure 18: Month of scarcity in the multi-area single-ORDC simulation.

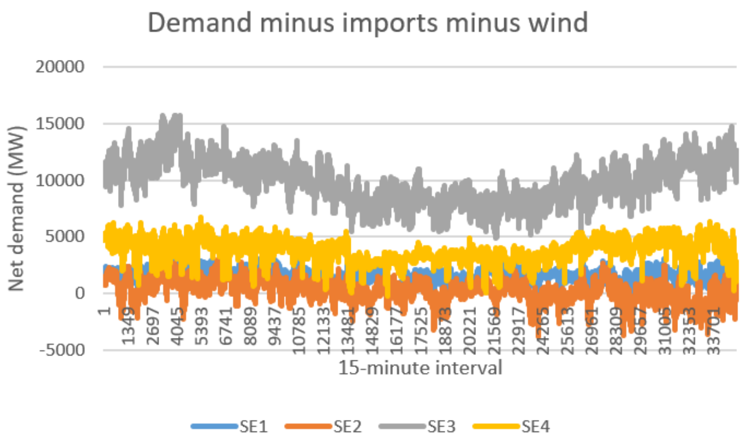

Figure 19: Net demand in the multi-area single-ORDC model, where we can observe that February corresponds to one of the most loaded periods for the system.

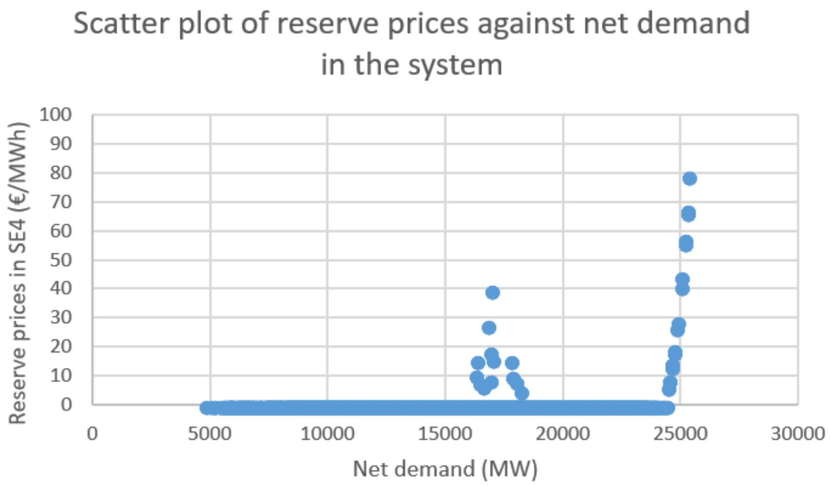

Figure 20: Scatter plot of reserve prices in SE4 versus net demand in Sweden. The hockey stick increase in the right of the graph corresponds to system-wide shortage in February. The bump in the middle of the graph corresponds to congestion in the SE3-SE4 link during August.

It may seem counter-intuitive that the system should be experiencing higher adders in February when it is not congested, than in August, when it is congested. Nevertheless, the scatter plot of Figure 20 offers an explanation for this: during February, when there is system-wide scarcity, the net demand is higher, which means that the system uses up more of the reserve stack, and thus less reserve demand is satisfied. The August scarcity events are scarcity events limited to the SE4 zone, and are not as severe as the system-wide scarcity events.

The fact that the periods of congestion correspond to scarcity of lower intensity, and produce zero reserve prices on the north side of the border, is confirmed in Figure 21. These periods correspond to the August scarcity incidents, and result in adders of lower intensity. Note that the reserve price is zero in the Northern part of the border. Even if BSPs in the North are offering reserve to the south, the excess of flexibility in the North implies that the price of reserve is 0 €/MWh in these zones.

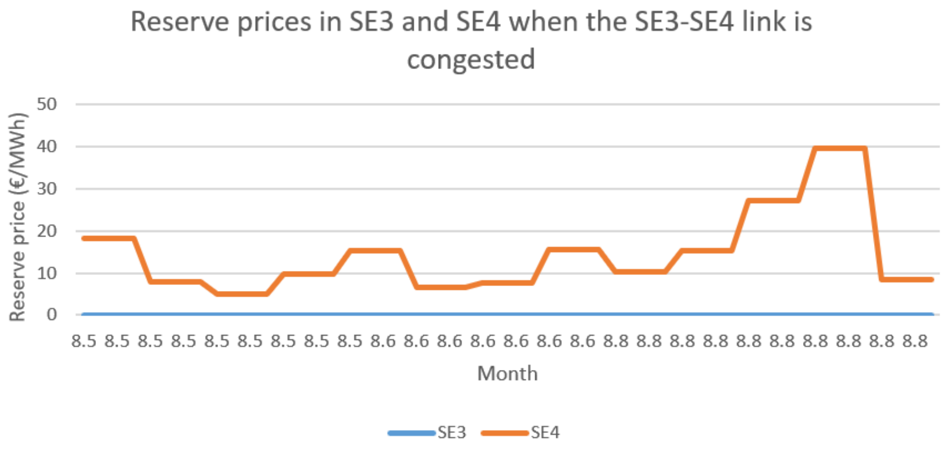

Figure 21: Reserve price in SE3 and SE4 during the periods when the SE3-SE4 link is congested.

The fact that the system is less likely to be congested during the winter periods is also corroborated by the fact that the thermal limits of lines are higher during the winter. This can be due to planned outages during the summer which imply that less transmission capacity is available in the summer, as well as lower ambient temperature during the winter which can increase the effective rating of the lines. This is confirmed by the input data of our study, which is presented in Figure 22, where we can indeed observe higher ratings for all lines during the winter months. Note that these are the total limits of the lines for both directions. The same figure includes plots of the ATC capacities, as they were provided in the original data sent to us by Svk. We note additionally that the maintenance of nuclear units is scheduled during the summer.

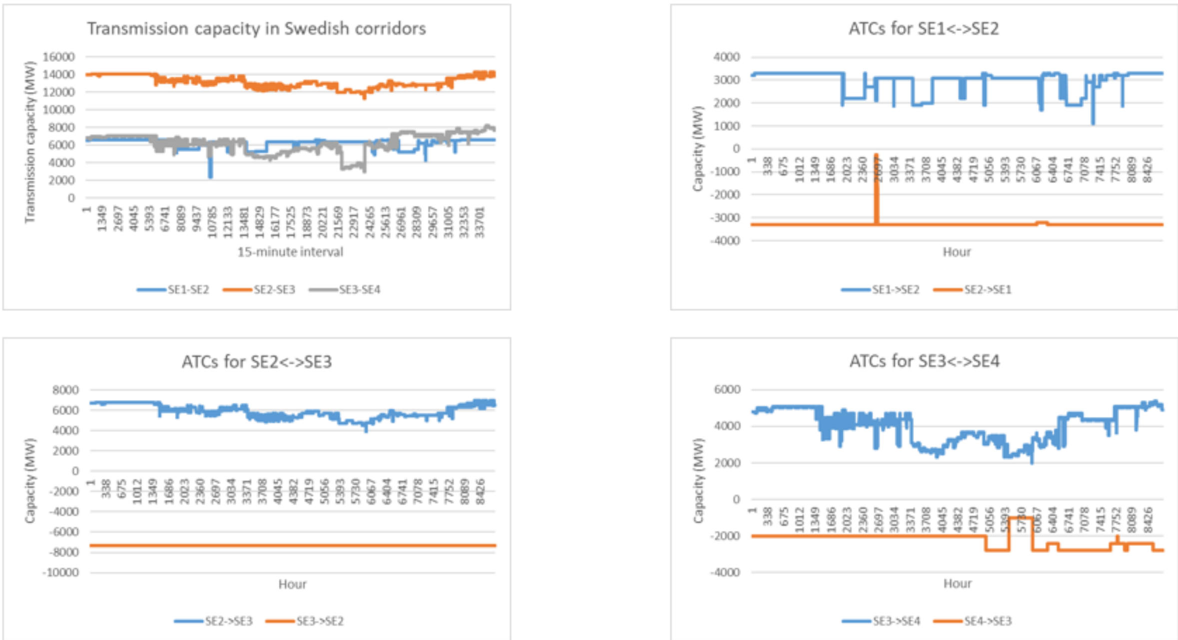

Figure 22: Transmission capacity along the Swedish corridors from January until December. Note a drop in the rating of all corridors during the summer, which may be attributable to scheduled maintenance of lines during the summer as well as increased ambient temperature during the winter.

3.4.2 Separate ORDCs in Each Swedish Zone

We now expand our model to account for multiple ORDCs in different zones. We specifically use the zonal imbalance. The statistics of the imbalance of each zone are summarized in Table 9. The resulting ORDCs are presented in Figure 23. The width of the ORDCs is driven by the standard deviation of each zone. For instance, as SE3 faces the greatest standard deviation, it has the widest ORDC.

Zone	Mean (MW)	Standard deviation (MW)
SE1	6.7	160.0
SE2	-28.3	256.1
SE3	-29.2	369.5
SE4	79.7	145.0

Table 9: Statistics of zonal balance in each Swedish zone which are used for deriving zonal ORDCs.

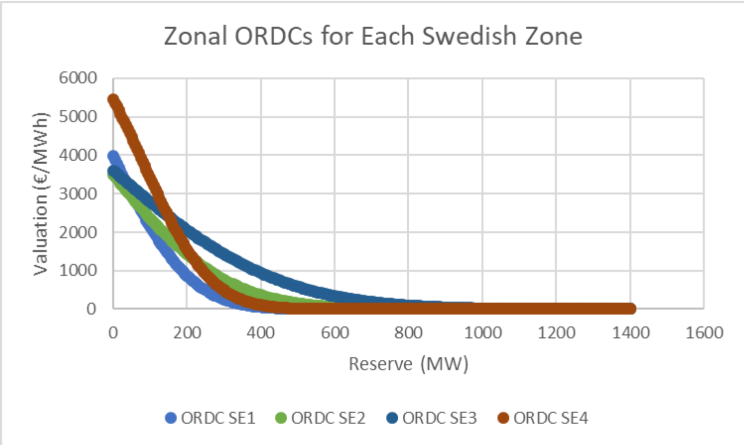

Figure 23: Zonal ORDC for the case study where each Swedish zone carries its own ORDC.

Figure 24 presents the energy stored in the hydro reservoir of each Swedish zone. As in the case of section 3.4.1, the initial and final level of hydro dams is presented in Table 8.

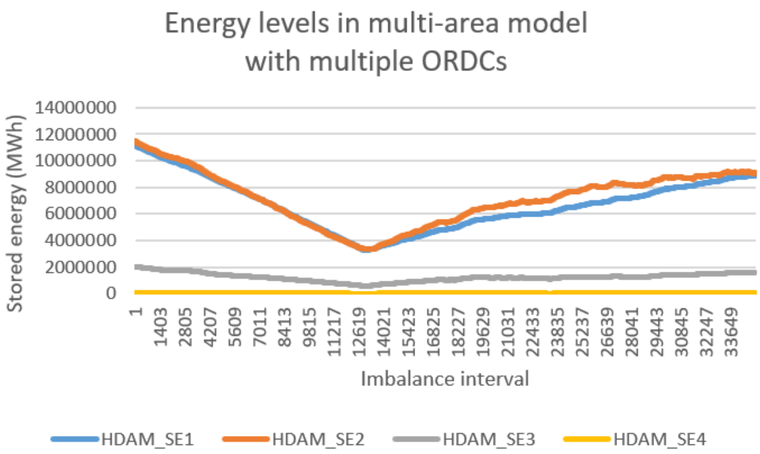

Figure 24: Stored energy in the hydro reservoirs of the multi-area model with multiple ORDCs.

The energy price is equal to 72.63 €/MWh in all zones. There are no periods during which load shedding occurs, and there is no congestion.

There are 1577 periods during which the scarcity adder is non-zero. The average value of the adder is 1.05 €/MWh. We plot the duration curve of reserve prices in Figure 25. Reserve prices are generally higher than in the case where only SE4 applies an ORDC.

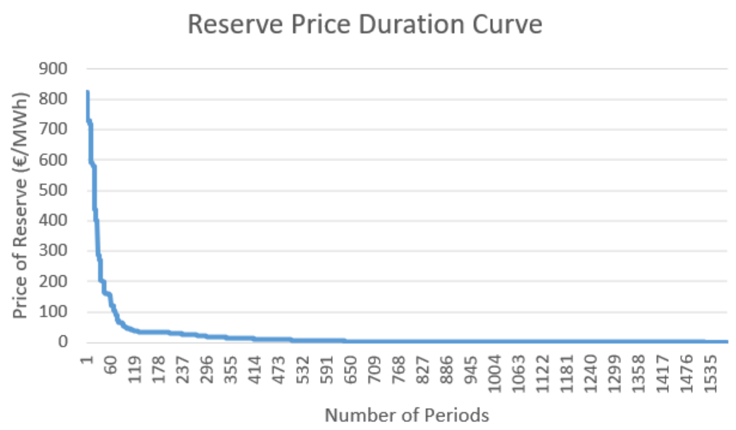

Figure 25: Duration curve of reserve price in the case where all zones apply an ORDC.

The precise timing of the scarcity incidents is presented in Figure 26. As in the case of the single-ORDC simulation, there is notable scarcity in February. This has already been explained, and is due to system-wide scarcity. There is, additionally, some scarcity observed in January and December, which are also highly loaded months, as indicated in Figure 19. The causality between net load in the system and the level of the adders is perfectly confirmed from the scatter plot of Figure 27, where we observe that net demand in the system is a reliable predictor of the level of scarcity adders.

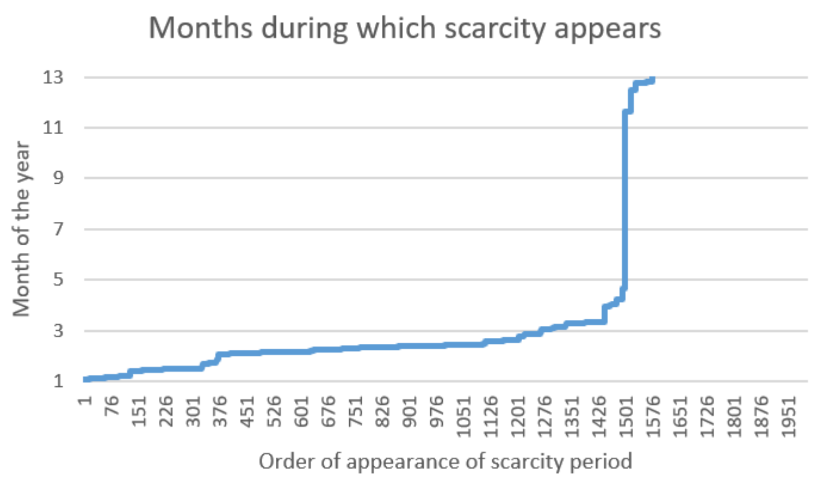

Figure 26: Months of scarcity in the multi-area single-ORDC simulation.

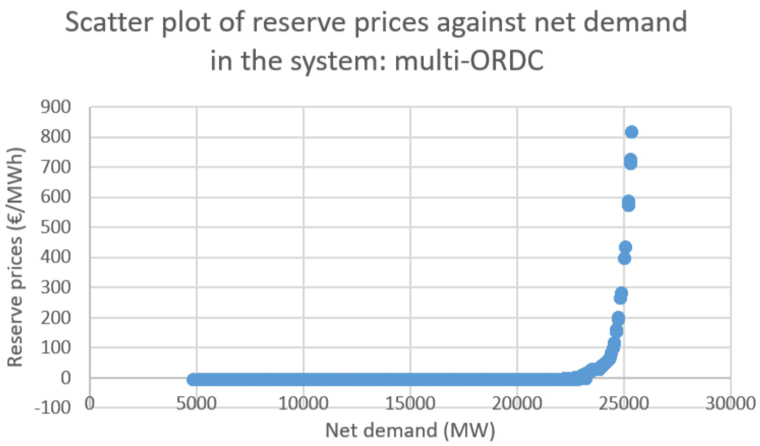

Figure 27: Scatter plot of reserve prices versus net demand in Sweden. The hockey stick increase in the right of the graph corresponds to system-wide shortage.

The fact that there is no congestion observed in the multi-ORDC case could be attributed to the fact that lower flows of energy and reserve can be expected from North to South. With ORDCs present also in the Northern zones, more of the potentially cheaper northern capacity is now precluded from producing energy, so as to cover local reserve requirements in the Northern zones, and also less of that reserve capacity is promised towards covering reserve requirements in SE4. The overall tendency, therefore, is for flows of energy and reserve towards the south to drop, and this is consistent with a reduction, or even elimination, of north-to-south congestion, but also an increase in the corresponding cost of operating the system and of system energy prices.

3.5 Removing Strategic Reserve

We discuss the distinction between strategic reserve and disturbance reserve in section 1.3 of the report. As discussed in section 1.3, an important market design choice in the context of scarcity pricing is whether or not these reserves are considered to contribute towards the scarcity adder or not. We approximate this design choice in this section by removing scarcity reserve capacity from the capacity that qualifies for scarcity pricing in the model (and from the model overall), while keeping all other aspects of the model of section 3.4 identical (same market model, and same distribution of ORDCs in the different zones). Svk has clarified that SE4 carried 560 MW of strategic period from November 15 until March 15 in 2021. We proceed to remove 560 MW of the condenser technology in SE4 from the model. Referring to table 7, this means that the condenser technology in SE4 is changed from 662 MW to 102 MW.

The energy levels of the reservoirs are not repeated, since they largely resemble the patterns of figure 15 and figure 24. The average price amounts to 74.67 €/MWh in all zones, which means that it increases relative to section 3.4, but without any congestion being introduced in the system. The average price of reserve is equal to 3.17 €/MWh, and is equal throughout the system.

There are 2234 periods during which the adder is non-zero (the criterion for non-zero adders is a scarcity price which is at least 0.01 €/MWh or greater). The price duration curve of the reserve price is presented in Figure 28.

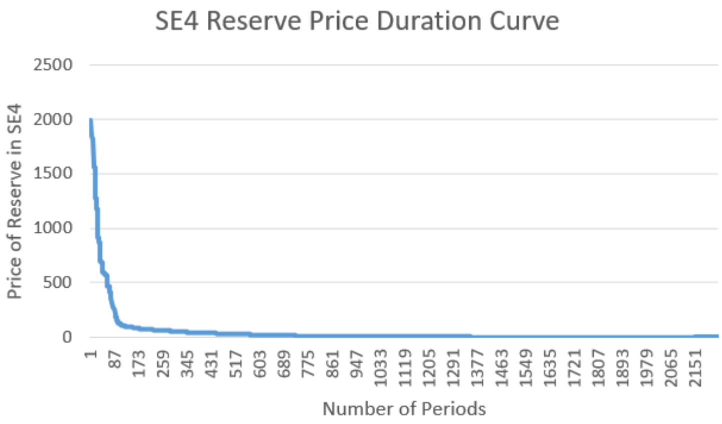

Figure 28: Duration curve of reserve price in the case where strategic reserve is removed.

The timing of the scarcity incidents confirms their connection to system-wide scarcity: as indicated in figure 29, the scarcity occurs during the months of peak demand, namely December, January and February.

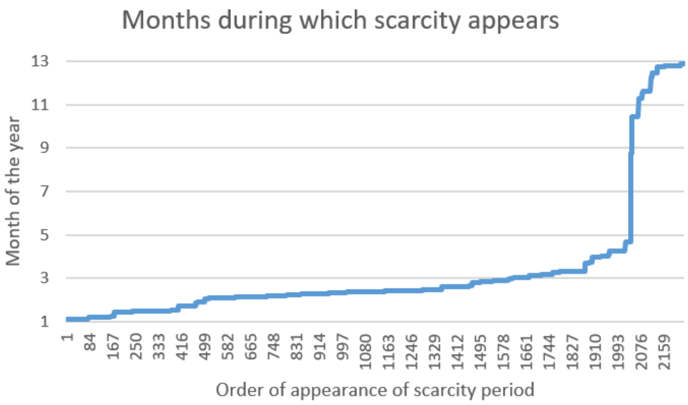

Figure 29: Months of scarcity in the simulation without strategic reserve.

The connection between system-wide scarcity and the non-zero reserve adders is confirmed by figure 30, where system net demand is plotted against the reserve adders. Again, it is confirmed that the scarcity pricing mechanism has direct connections with adequacy, by providing signals when the capacity of the system is tight.

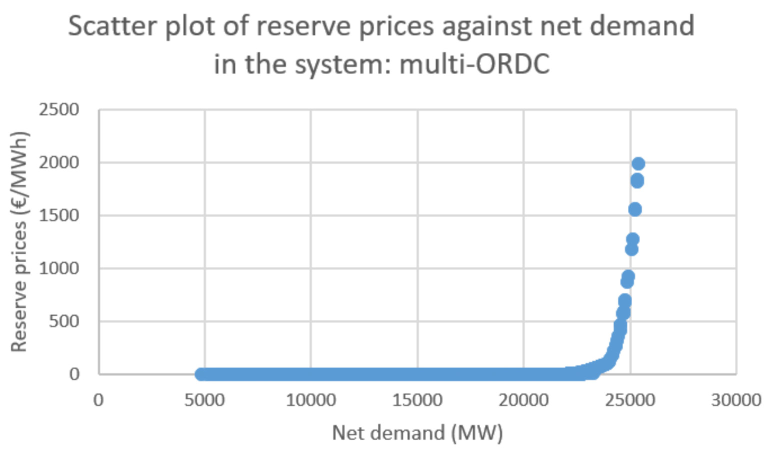

Figure 30: Scatter plot of reserve prices versus net demand in Sweden in the case without strategic reserve. The hockey stick increase in the right of the graph corresponds to system-wide shortage.

3.6 Removing Strategic and Disturbance Reserve

In the same spirit as in section 3.5, and motivated by the discussion of section 1.3, we consider a run of the model, where both strategic reserve and disturbance reserve are removed from the supply stack. It has been indicated to us by Svk that disturbance reserve corresponds to 764 MW of capacity in SE3 and 540 MW of capacity in SE4. This has been removed from the available OCGT capacity in table 7. This in turn implies that the OCGT capacity in SE3 has been reduced from 962 MW to 198 MW, whereas the OCGT capacity in SE4 has been reduced from 618 MW to 78 MW. The reductions in strategic reserve have also been inherited identically from section 3.5.

The energy levels of the reservoirs are not repeated, since they largely resemble the patterns of figure 15 and figure 24. The average price amounts to 4830.3 €/MWh in all zones. Given the degree of scarcity that is introduced in the system, it becomes clear that a viable implementation of the mechanism should not exclude both disturbance as well as strategic reserve capacity from the estimated available flexible capacity, especially if these technologies are technically capable of delivering fast-responding real-time energy to system operation.

4. Where to Apply the Adders, Flexibility Incentives and Back-Propagation

4.1 Back-Propagation and the Law of One Price

Having introduced scarcity pricing adders as an approximation of the outcome of a co-optimization in energy and reserves, we now proceed to discuss where these adders should be applied in order to approximate the outcome of an energy and reserve co-optimization as closely as possible. In particular, and given the separation of balancing and imbalance settlement as well as the absence of a real-time market for reserve in European balancing markets (see Table 2 and [12]) the question that has emerged repeatedly in the context of stakeholder engagement for the implementation of scarcity pricing is where these scarcity adders should be applied, namely:

- Should the scarcity adders uplift the balancing price that applies to BSPs?
- Should the scarcity adders uplift imbalance settlement that applies to BRPs?
- Should the scarcity adders be used for settling real-time reserve imbalances (i.e. for implementing a real-time market for reserve capacity, in the same way as we have energy imbalance settlement)?

The proposal in Papavasiliou [12] advocates that “scarcity adders should apply to all of the above”. In this section we introduce two fundamental concepts that form the basis of the proposal of [12] for the implementation of scarcity pricing in European balancing markets: *back-propagation* and the *law of one price*.

Back-propagation is a fundamental concept in economics, and in particular in financial markets [13, 15]. The idea of back-propagation is that, in a risk-neutral environment, forward markets for an underlying commodity or service tend to align around the expected real-time price of the commodity. One can see why this happens by reasoning in terms of contradiction. Suppose that average forward (e.g. day-ahead) prices for a commodity or service are higher than the average real-time prices for this commodity or service. Then buyers of the commodity will refuse to buy in the day ahead, and will rather wait to buy until real time. Similarly, sellers will try to underbid each other in order to sell in the day-ahead, since it is more favorable than trading in real time. The net effect of the withdrawal of demand and the more intense competition of suppliers in the day ahead is that day-ahead prices will drop. Prices will continue to drop until the day-ahead market price aligns with the average real-time price. If the average real-time price is lower than the day-ahead price, the exact opposite process takes place, but the end effect is the same: forward (day-ahead) prices tend to equalize to average real-time prices.

The **law of one price** [14] (stated originally in 1879 by Jevons) is another fundamental concept in economics. It stipulates that a unique product or service should be remunerate at a unique price. The idea is that if multiple prices exist for the same product or service in the market, then suppliers will opt for the higher price, buyers will opt for the lower price. This will exert downward and upward pressure on the higher and lower price respectively, until an equilibrium is reached where said product or service is traded at a unique price uniformly throughout the market.

What happens if there is no real-time market for a service or commodity (e.g. reserve), i.e. if the service or commodity is given away for free in real time? Then one can expect that the forward price of said commodity amounts to zero. And what happens if those who need balancing energy (e.g. BRPs) are settled at a different price than those who provide it (e.g. BSPs)? Then one can expect that BSPs will migrate to the price that best rewards them (if needed by self-dispatching in real time), until imbalance settlement aligns with balancing settlement[12].

This simple reasoning is the basis for a quantitative development which justifies our proposal for the coherent implementation of scarcity pricing [12]. Apart from justifying our own proposal, we use our quantitative methodology in order to reason about the equilibrium that one could expect from various alternative designs that have been proposed in the context of the implementation of scarcity pricing in European balancing markets [12]. Although we do not enter the full detail of the development in [12] here, we nevertheless highlight the key arguments and conclusions of the analysis.

It is worth noting that our analysis stylizes various real-world aspects of the problem. For instance, (1) certain resources (such as certain categories of demand response) may be able to resort to self-balancing but may not qualify for certain balancing products. (2) Balancing prices may be predictable to a certain extent. (3) The provision of reserve may involve a certain cost of commitment. (4) Multiple reserve products exist in the real market, and not all of them are activated on the basis of pure economic signals. To these potential misalignments between theory and the real world we provide two counter-arguments: (i) A sound market design should pass the test of the simplest possible setting first. If incentives fail to align under idealized settings, then this indicates fundamental flaws of a given market design. It would be wishful thinking to expect that the complexities of a real-world market would somehow “correct” a market design proposal which fails to deliver under idealized simplifying assumptions.

---

[12] Note that different TSOs have different views on this matter. For instance, ELIA, for instance, has expressed a strong position against unbalanced portfolios in the past, although nowadays the organization seems very open to self-dispatch. Tennet, as another example, is known to be open to self-dispatch.

Notwithstanding these various view on self-dispatch, there exists a consensus that it is difficult for the TSO to enforce the requirement of balanced portfolios. For instance, portfolio owners can easily override this requirement by misrepresenting renewable forecasts to the TSO, due to information asymmetry, thereby replicating the effect of an unbalanced portfolio. This raises an interesting question of whether publishing prices in real time can induce portfolios to react proactively.

Svk indicates a preference to have balanced portfolios. According to Svk, unbalanced portfolios are easier to cope with when a TSO is not facing bottlenecks within its bidding zones. Svk indicates that, in a system with potential congestions, the TSO can face problems if portfolios start reacting to imbalances, since this can lead to congestions between bidding zones. The future move to a 15-minute ISP may make proactive system balancing by portfolios more acceptable to Svk.

In Svk there is an imbalance fee of 1.15 €/MWh for covering TSO costs, which is harmonized with the rest of the Nordics (except Energinet, which applies a different imbalance fee).

We finally mention that there is a tendency for portfolios to be long positioned in the Nordics, based on empirical observations provided by Nordic TSOs during the project.

(ii) Our quantitative framework provides the conceptual framework for reasoning about generalizations which can handle some of the above real-world complexities. These generalizations are part of ongoing work by the group of Professor Papavasiliou, and part of academic research within his team, but fall out of the scope of the current Svk contract. Nevertheless, any insights that emerge / are published from this research can immediately be integrated in the Svk deliverable. (iii) The fact that we cannot prove rigorous results under certain complex real-world assumptions for certain market design proposals does not justify speculations about the performance of alternative design proposals which deviate from first principles.

4.2 Quantitative Framework

We consider a generic agent which owns a portfolio of both flexible as well as inflexible assets, and is active in a market for a single reserve product. The agent is thus exposed to a possible imbalance in real time, and may choose to activate its flexible assets for the provision of balancing energy. In addition to these real-time decisions, the agent may choose to bid in the day-ahead energy and reserve market. The sequence of events that the agent is facing in its decision making are depicted in Figure 14. Concretely:

- <u>First square</u>: In the day ahead, the agent decides on how to bid in the day-ahead reserve market.
- <u>First oval</u>: Based on the bids collected from all agents, it is determined whether each agent is cleared or not for the provision of balancing capacity. If the agent is cleared for balancing capacity, it is obliged to bid at least the amount of balancing capacity for which it was cleared in the balancing market.
- <u>Second square</u>: Given the outcome of the balancing market, and after observing the imbalance within its own portfolio, each agent decides whether to activate its flexible capacity within its own portfolio, or whether to bid (part of) its capacity to the balancing market. The outcome of the day-ahead reserve market constrains this decision to a certain extent: As we mention earlier, if the agent is cleared in the day ahead for providing reserve, at least the cleared amount of balancing capacity must be bid into the balancing market as balancing energy.
- <u>Second oval</u>: Given the realization of imbalances within other portfolios and the market at large, the system imbalance is revealed, and the TSO activates available balancing resources. The activation of balancing resources determines the balancing market price, and this also implies imbalance settlement (which differs among different market designs, as we discuss below).

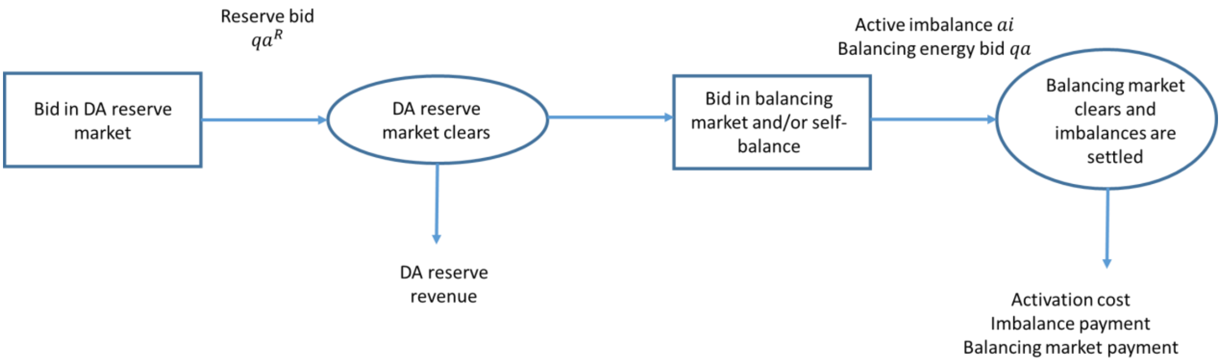

Figure 28: Sequence of events in our analytical framework. Squares represent decisions , ovals represent realizations of random outcomes and corresponding payoffs.

Note that the quantitative framework that is described in Figure 28 follows the structure of a Markov decision process (MDP) [16], which is a quantitative framework for multi-stage decision making under uncertainty. This implies that the analytical results which we present in the sequel can also be validated by simulation through algorithms that can be used for solving MDPs. This experimental validation is exactly undertaken in [17], where we confirm experimentally all the theoretical results that our theory anticipates.

Note, also, that the framework described previously is one that can be used for computing a market equilibrium. Concretely, by analyzing the optimal decisions of a so-called fringe agent (i.e. an agent whose size is too small to affect the market outcome), we are able to derive the optimal strategy of this agent. The overall market is depicted graphically in Figure 29. We can then check whether this optimal strategy is consistent with what the agent assumes would be the optimal strategy of its competitors when it chooses its own optimal strategy. In other words, we not only characterize the optimal strategy of a given fringe agent, but also characterize a **pure-strategy Nash equilibrium** of the market.

As we mention previously, the results that we present next are guaranteed under the framework described above. We do not make claims about alternative settings, and the analysis of such settings falls beyond the scope of the existing work.

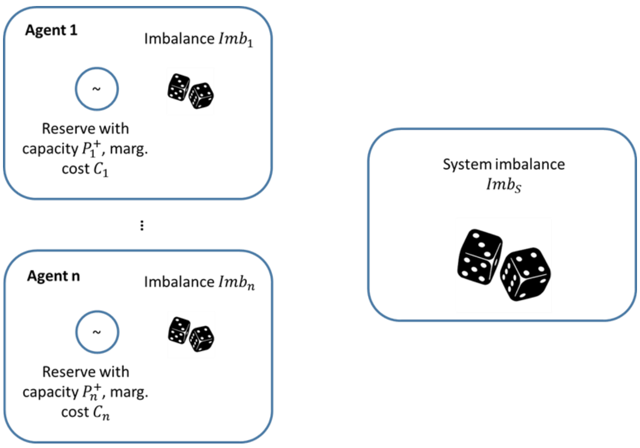

Figure 29: Graphical depiction of overall market in the quantitative analysis of [12].

4.3 Existing European Balancing Market Design

We consider a “vanilla” European balancing market design, whereby agents are paid a balancing price for activated balancing energy, are settled for imbalances with respect to the part of their portfolio that is not bid into the balancing market, and are paid a uniform price in the day-ahead reserve auction. In other words, the payoff of an agent in real time can be described as follows:

$$\lambda^B \cdot qa - \lambda^I \cdot (Imb - ai) - C \cdot (qa + ai) \qquad (4.3.1)$$

The notation here is as follows:
λ^B: The balancing price
λ^I: The imbalance price
C: The marginal cost of the fringe agent that we are considering
qa: The amount of reserve capacity that is activated for balancing by the system operator.
Imb: The imbalance that the agent under consideration faces within its portfolio.
ai: An active imbalance that the agent may choose to effect on its portfolio by dispatching its reserve asset upward or downward, without being asked to do so by the system operator. Positive ai corresponds to an upward activation, i.e. a negative contribution to imbalance.

We are interested in characterizing the optimal bidding strategy of an agent in this setting. The detailed mathematical development is provided in [17], its technical appendix [18], as well as [19] (where additional market designs are analyzed). Here we summarize the key insights of the analysis.

The first thing to observe is that an agent has an interest in bidding its marginal cost truthfully in the balancing market. Overstating its cost would increase its chances of not being profitably activated when the system is short, whereas understating its cost would increase its chances of being activated at a financial loss.

The next thing to note is that, when deciding whether or not to self-balancing its portfolio (by choosing a non-zero active imbalance ai), the agent is trading off the potential profits from the activation of its balancing capacity in the balancing market against the benefits of recuperating an imbalance price. When balancing prices and imbalance prices are aligned ($\lambda^B = \lambda^I$), the latter option is clearly not advantageous, because the agent will sometimes be paid an imbalance price below its marginal cost, whereas if the agent bids truthfully in the balancing auction it is only activated when the balancing price is profitable for the agent.

This is a positive result: when the balancing price is equal to the imbalance price, $\lambda^B = \lambda^I$, a fringe agent has an interest to bid its entire capacity to the balancing auction at its true marginal cost. This maximizes the degrees of freedom of the system operator, and promotes price discovery and operational efficiency in real time.

Perhaps counterintuitively, this virtuous attribute of the “vanilla design” is, at the same time, what undermines the ability of the design to back-propagate the real-time value of reserve. Concretely, since the agent has an interest in bidding its balancing capacity in the balancing auction anyway, it has no opportunity cost associated to offering this reserve capacity in the day-ahead reserve market. This latter effect undermines the back-propagation of scarcity prices based on ORDC, and it is exactly this problem that is corrected by putting in place a real-time market for reserve capacity.

With that being said, we know that existing day-ahead reserve markets *do* have a non-zero clearing price. This may be due to a number of factors that are not accounted for in the present analysis. For instance, in our analysis we do not account for the cost of committing resources so as to make reserve available in the first place. We also consider fringe agents that are not able to affect market prices through strategic behavior. The fact that, however, is that a desirable design should be able to deliver the desirable attribute of back-propagation under idealized conditions, even if these fail to hold perfectly in practice. Failure to do so indicates potentially fundamental pitfalls in the market design.

4.4 A Disciplined Approximation of Co-Optimization

A disciplined approximation of the principles that are implied by the co-optimization model of section 3.4 is based on two observations: (1) there is a unique price for real-time energy in the model (which implements the law of one price for energy in real time), and (2) the model also implements a real-time market for reserve capacity. Adapting this framework to European balancing markets implies the following payoffs to a fringe agent in real time:

$$\begin{aligned}(\lambda^B + \lambda^R) \cdot qa - (\lambda^B + \lambda^R) \cdot (Imb - ai) - C \cdot (qa + ai) \\ + \lambda^R \cdot (P^+ - qa - ai) - \lambda^R \cdot qa^R\end{aligned} \quad (4.4.1)$$

The following additional notation is introduced in this context:
λ^R: The scarcity adder.
P^+: The flexible capacity of the fringe agent.
qa^R: The reserve capacity that is sold by the fringe agent in the day-ahead reserve market.

The additional settlement components relative to equation (4.3.1) are highlighted in colored text. Concretely:

- The blue term represents the application of the scarcity adder as an add-on to the balancing price
- The orange term represents the application of the scarcity adder as an add-on to the balancing price for the purpose of imbalance settlement (this effectively implements the law of one price in real-time energy)
- The green term implements a real-time market for reserve, i.e. a market where reserve imbalances are settled against the market price of reserve, which is precisely the value of the scarcity adder. On the one hand, the agent is paid a real-time price for reserve for any amount of reserve capacity that it makes available in real time. On the other hand, the agent must buy back the amount of reserve capacity that it has sold to the system operator in the day-ahead market at the real-time price of reserve. It is exactly the presence of this last

term in the settlement that permits the back-propagation of scarcity prices to the day-ahead reserve market.

The quantitative analysis of [17-19] proves mathematically that this design safeguards the truthful bidding of the full capacity of the agent to the balancing market without compromising the back-propagation of the value of reserve to the day-ahead reserve market. Concretely, since the agent now sells real-time reserve at a real-time reserve price λ^R, its opportunity cost for selling this reserve in the day-ahead reserve market becomes the average real-time value of reserve *due to back-propagation*.

In terms of financial neutrality, note that the cash flows implied by this proposal are financially neutral in the balancing energy market, since the adder applies to both balancing prices as well as imbalance settlement. On the other hand, the TSO acts as a single buyer in the real-time balancing capacity market. As such, it procures reliability as a public good on behalf of the system, and charges it back to BRPs, in exactly the same spirit as the day-ahead reserve market. As a market for residual imbalances, the real-time market for reserve may likely involve the trading of limited volumes, thus the procurement costs of the TSO may be expected to be lower than those of the day-ahead market for reserve.

ELIA has challenged the alignment of balancing prices and imbalance settlement [21]. Concretely, it is noted that:

"*[…] the equalization of the imbalance tariff to the uniform price of activated balancing energy bids is considered not compatible with the prevailing market design. At fundamental level, the BSP is given a European-wide role, while the role of the BRP remains crucially associated to the LFC area. Therefore, the imbalance tariff that applies to the BRP depends in the first place on the situation within its LFC area, while the balancing energy price that applies to the BSP results from a European-wide optimization, combining the requests from all LFC areas and taking into account possible congestions. This means that the imbalance tariff cannot be guaranteed to be equal to the uniform price of activated balancing energy bids, as the necessary activations may differ per LFC area.*"

To support this argument, the following hypothetical example is provided in [21], which is depicted in Figure 30:

"*Consider for instance the following example. Assume that the Belgian and French LFC areas form an uncongested area. Belgium is 100 MW long, while France is 500 MW short. In the uncongested area therefore, 400 MW upward mFRR will be activated. However, the imbalance tariff in the Belgian LFC area should not be driven by such activation as no upward mFRR has been activated to cover the demand of the Belgian TSO. In such case, assuming no aFRR has been activated in Belgium either, the imbalance tariff will be based on the Value of Avoided Activation, in accordance with art. 10 of the Imbalance Settlement Harmonization methodology. Setting the Belgian imbalance tariff at the price of upward mFRR would incentivize Belgian BRPs to take a long position, aggravating the Belgian LFC Block imbalance. In addition, would such behavior occur, this would not only cause much more volatile system imbalances, it would also increase Belgian's reserve requirements. Indeed,* *the dimensioning of reserves is function of the system imbalances, and henceforth of the quality of the system balance.*"

Figure 30: The example described in page 42 of [21] for challenging the alignment of balancing and imbalance settlement.

The argument and example presented by ELIA appear to raise a cross-border concern, which is the focus of material that comes later in the present report. The example seems to be missing some essential information, and does not identify what is deemed the optimal outcome in the setting of the example above. For instance, without a description of the marginal costs of the various balancing resources in the two zones, it is not possible to determine the economically efficient outcome of this cross-border example.

It has been pointed out by the Nordic TSOs that reserves are anyway shared in the Nordics, but that pricing can be discussed for the future Nordic Balancing Model. In the hypothetical example of Figure 30, the Nordic TSOs point out that they would expect both areas to have the same imbalance price (which is coherent with what a co-optimization result would produce if the line is not congested). Currently, it is expected by Nordic TSOs that the imbalance price of an uncongested area should be the same throughout the area, but this may change with the deployment of MARI and PICASSO. The preference that has been indicated by Nordic TSOs is to have a forward-looking view in this project, and align the design proposed in this project with what would be put in place in anticipation of MARI and PICASSO. The hope is that a common imbalance settlement will be agreed throughout the entire Nordic region.

4.5 Applying Adders on Imbalances Only

An alternative proposal that has been advocated by certain market stakeholders [20, 21] is to apply the scarcity adder on imbalance settlement only. Although this proposal is not backed up by quantitative analysis in [20, 21], a detailed mathematical analysis of the proposed design is provided in [17-19]. We summarize the conclusions of this analysis in the present section.

The precise real-time settlements that would apply under this design can be described as follows:

$$\lambda^B \cdot qa - (\lambda^B + \lambda^R) \cdot (Imb - ai) - C \cdot (qa + ai) \qquad (4.5.1)$$

Note that only the orange term from equation (4.4.1) is retained.

Under this design, the tradeoff of agents to use their balancing capacity in order to self-balance their portfolio now creates conditions for certain agents (those with the lowest marginal costs in the merit order stack for upward activation) to take their chances by activating their reserve upwards without being asked to do so by the system operator. This is indicated graphically in Figure 31. With the presence of an ORDC adder, the payoff of the imbalance price can be adequate to cover the losses of sometimes self-dispatching with an imbalance price which is lower than the marginal cost of the resource which is used for self-balancing. Since the balancing market does not apply the same adder, this will make it advantageous for cheaper resources in the upward balancing stack to migrate towards self-balancing through active imbalances of their reserve resources, and keeping these resources out of the balancing market.

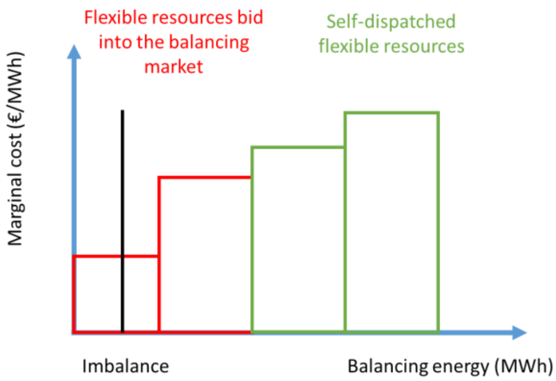

Figure 31: Merit order curve under the design where scarcity adders are only applied to BRPs. Red resources choose to self-dispatch in order to capture the scarcity adder. Green resources choose to bid in the balancing market. When imbalances are not too large, the self-dispatch of the red resources becomes inefficient.

This effect is interesting, because it creates an opportunity cost for bidding these resources into the balancing market, and therefore produces a back-propagation of a day-ahead reserve price. However, the opportunity cost is lower than that of the average ORDC adder [18], and only applies to a subset of the balancing resources. Therefore, something is being back-propagated, but it is not the average ORDC scarcity adder. Moreover, it is worth noting that active imbalances result in economic inefficiencies, since the imbalances that are being resolved within a balancing portfolio could have been balanced by potentially cheaper resources in the balancing market.

Another interesting effect of this design is the fact that the resources which are self-balancing are replacing resources that would have otherwise been activated in the balancing market. This depresses the balancing energy price, and counteracts the effect that the scarcity adder has on the imbalance price. Alas, the fundamental problem remains [18]: in the absence of a real-time market for reserve capacity, back-propagation is undermined.

The fact that this design proposal induces self-dispatch can be problematic in systems with significant internal congestion. Although the issue of internal congestion is not entirely pertinent in systems like Belgium where this design has been debated more extensively, one can foresee more significant challenges resulting from self-dispatch in systems with internal congestion.

The fact that this design induces low-marginal-cost agents to self-dispatch upward can produce inefficient dispatch, because there are certain periods when these agents are activated upwards when this is not warranted. This would then result in a counter-activation by cheaper agents that are available for downward balancing. This results in an overall inefficiency, whereby upward self-dispatch is counter-balanced by downward dispatch in the balancing market, but the resources that are dispatched downward are cheaper than the resources that are self-dispatching upward. This aspect of the design can be problematic, as one could argue that it contravenes article 3(m) of the Clean Energy Package [22]:

"*Member States, regulatory authorities, transmission system operators, distribution system operators, market operators and delegated operators shall ensure that electricity markets are operated in accordance with the following principles:* ***market rules shall enable the efficient dispatch of generation assets, energy storage and demand response****;*"

The insights of the analysis presented above are summarized in Table 10.

	Adder on balancing price?	Adder on imbalance settlement?	Real-time market for reserve?	Back-propagation of real-time value of reserve?	Bid flexibility in balancing market?
"Vanilla" EU balancing market	No	No	No	No	Yes
Disciplined approximation of co-optimization	Yes	Yes	Yes	Yes	Yes
Adder on imbalance settlement only	No	Yes	No	Weak	Not always

Table 10: A summary of the pros and cons of alternative design options for implementing scarcity pricing.

5. Implementation

In this section we address various topics that relate to the practical implementation of scarcity pricing. We focus on three specific aspects: legal and institutional compliance, the products that are affected, and the compatibility of the mechanism with CRMs.

5.1 Legal and Institutional Compliance

The Clean Energy Package (CEP) of the European Parliament [22] and the Electricity Balancing Guideline (EBGL) of the European Commission [23] are milestone documents insofar as the implementation of scarcity pricing in the European market design is concerned.

The CEP highlights the importance of scarcity pricing in whereas (23) of the CEP [22]:

"*While decarbonisation of the electricity sector, with energy from renewable sources becoming a major part of the market, is one of the goals of the Energy Union, it is crucial that the market removes existing barriers to cross-border trade and encourages investments into supporting infrastructure, for example, more flexible generation, interconnection, demand response and energy storage. To support this shift to variable and distributed generation, and to ensure that energy market principles are the basis for the Union's electricity markets of the future, a renewed focus on short-term markets and scarcity pricing is essential.*"

Whereas (24) of the CEP [22] underlines the pay-for-performance attributes of scarcity pricing which we analyze in section 1, and further touches on the interaction between scarcity pricing and capacity mechanisms:

"*Short-term markets improve liquidity and competition by enabling more resources to participate fully in the market, especially those resources that are more flexible. Effective scarcity pricing will encourage market participants to react to market signals and to be available when the market most needs them and ensures that they can recover their costs in the wholesale market. It is therefore critical to ensure that administrative and implicit price caps are removed in order to allow for scarcity pricing. When fully embedded in the market structure, short-term markets and scarcity pricing contribute to the removal of other market distortive measures, such as capacity mechanisms, in order to ensure security of supply.*"

Whereas (24) specifically makes reference to the ability of scarcity pricing to mitigate missing money issues, and remarks on how this interplays with capacity markets. We provide a formal quantitative analysis of this phenomenon in section 5.3.

The possibility of a unilateral implementation of a scarcity pricing mechanism by a Member State is foreseen by Regulation 2019/943 of the European Parliament [22]. Concretely, it is mentioned in article 20(3) of Regulation 2019/943 [22]:

"*Member States with identified resource adequacy concerns shall develop and publish an implementation plan with a timeline for adopting measures to eliminate any identified regulatory distortions or market failures as a part of the State aid process. 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 2017/2195;"

The possibility of a unilateral implementation is further indicated in the EBGL [23]. Note that article 5.4(c) of EBGL [23] reads as follows:

"*The proposals for the following terms and conditions or methodologies shall be subject to approval by each regulatory authority of each concerned Member State on a case-by-case basis: the terms and conditions related to balancing pursuant to **Article 18**;*"

Reading on to article 18 of [23], we find in article 18.5(i) of [23]:

"*The terms and conditions for **balancing service providers** shall contain: the rules for the settlement of balancing service providers defined pursuant to Chapters 2 and 5 of Title V;*"

And in article 18.6(k) of [23]:

"*The terms and conditions for **balance responsible parties** shall contain: the settlement rules pursuant to Articles 52, 53, 54 and 55;*"

The EBGL thus indicates that the definition of the settlement rules for balancing and imbalance pricing can be decided unilaterally in each Member State.

An important question that has emerged among stakeholders is the scope of applicability of a scarcity pricing function, i.e. whether a scarcity adder can apply to BSPs, BRPs, or both. This is directly relevant to the discussion of market design that is developed in section 4, because the legal provisions set the boundary conditions of what is feasible in the practical implementation of scarcity pricing.

The application of a scarcity adder to imbalance settlement applicable to BRPs is not debated among stakeholders. TSOs have considerable control over imbalance settlements, and there are various examples in European markets whereby imbalance settlement is modified relative to balancing market prices. For instance, the Nordics (with the exception of Energinet) apply an imbalance charge of 1.15 €/MWh for covering TSO costs. Belgium applies a so-called alpha component [5] which is a surcharge on balancing market prices when the system is very long or very short, and is intended to discourage BRPs from being out of balance when the Belgian area is highly stressed. In a similar spirit, one can consider the introduction of a scarcity component to imbalance settlement, and the basis of doing so does not appear to encounter legal obstacles [12].

This view is further reinforced by article 44(3) of the EBGL [23]:

"Each TSO may develop a proposal for an additional settlement mechanism separate from the imbalance settlement, to settle the procurement costs of balancing capacity pursuant to Chapter 5 of this Title, administrative costs and other costs related to balancing. The additional settlement mechanism shall apply to balance responsible parties. ***This should be preferably achieved with the introduction of a shortage pricing function****. If TSOs choose another mechanism, they should justify this in the proposal. Such a proposal shall be subject to approval by the relevant regulatory authority."*

ELIA has challenged the legal basis for the application of scarcity adders to BSPs. Concretely, ELIA argues that article 20(3) of the CEP [22], which refers to article 44(3) of the EBGL [23], limits the scope of scarcity pricing only to BRPs.

The CREG has developed the counterargument [6] that article 20(3) of the CEP targets BSPs. The basis of the argument of the CREG is (i) the wording of article 20, (ii) the position of ACER, and (iii) the position of the European Commission on the matter. We cover these three points in turn.

Regarding the wording in article 20(3), even if the article makes reference to article 44(3) of the EBGL, it still refers to balancing energy. Balancing energy and balancing capacity are defined[13] in articles 2.11 and 2.13 of the CEP [22], and appear to target BSPs.

The CREG further points to annex I of ACER Decision 01/2020 [30]. According to the CREG, article 1(4) of [30], indicates that article 20(3) of the CEP [22] and article 18 of the EBGL [23] target BSPs:

"*This pricing methodology is without prejudice to the introduction of a* ***shortage pricing function for balancing energy*** *as referred in* ***Article 20(3) of the Regulation (EU) 2019/943****, within the national terms and conditions related to balancing pursuant to* ***article 18 of the EB Regulation****.*"

The CREG further points to the response of the European Commission [36] to the Implementation Plan of the Belgian government [35] as a further basis for establishing that scarcity adders cover BSP settlements, in addition to BRP settlements:

"***The Commission, however, invites Belgium consider whether the scarcity pricing function should apply not only to BRPs but also to balancing service providers (BSPs)****. This may support security of supply by* ***ensuring that BRPs and BSPs face the same price for the energy produced/consumed, as price differentiation may result in inefficient arbitrage from market players****. The Commission also considers that the scarcity pricing function should be triggered by*

[13] Article 2.11: "*'balancing energy' means energy used by transmission system operators to carry out balancing;*".
Article 2.13: "*'balancing capacity' means a volume of capacity that a balancing service provider has agreed to hold and in respect to which the balancing service provider has agreed to submit bids for a corresponding volume of balancing energy to the transmission system operator for the duration of the contract;*".

the scarcity of reserves in the system and it should be calibrated to increase balancing energy prices to the Value of Lost Load when the system runs out of reserves."

The position of the European Commission thus seems to support not only the principle of applying adders to BSPs, but also the law of one price invoked earlier in this section.

Assuming that one accepts the interpretation of the CREG as being correct (i.e. that the CEP and EBGL allow the application of a scarcity pricing function to BSPs as well as BRPs), then it can be argued that this opens the path towards also implementing the settlement of a real-time market for reserve (i.e. the green term in equation 4.4.1).

ELIA has further raised concerns about whether scarcity pricing constitutes State Aid. For State Aid to exist, four cumulative conditions must be fulfilled [37]: (i) the measure must 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 position of the CREG regarding State Aid is that condition 2 does not hold. In the view of the CREG, 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 / activation . Moreover, the European Commission itself encourages Belgium to consider implementing the mechanism [36], thus it would appear paradoxical to encourage the implementation of the mechanism only to then rule that it constitutes State Aid. Finally, since scarcity pricing implements a real-time market for reserve, considering this as State Aid could likely also raise questions of whether the day-ahead market for reserve also constitutes State Aid.

We finally note that in the CRM State Aid reviews conducted by the European Commission [37, 38, 32] the text supports or considers the implementation or improvement of scarcity pricing mechanisms in the scope of the implementation of article 20 of the CEP. For instance, we read in recitals 16(e), 149 and 156 of the Polish CRM [38]:

"*By 1 January 2021, **Poland will introduce an administrative scarcity pricing mechanism** as referred to in Article 44(3) of the Electricity Balancing Guideline. The mechanism will be designed to provide a price adder to the energy prices on the balancing market varying in function of the amount of the reserve margin in the Polish system. The price adder calculation will be based on the Value of Lost Load (VoLL) and the Loss of Load Probability (LoLP), ensuring that when reserves are exhausted (i.e. there are no more available reserves that can be activated by the TSO) the imbalance settlement prices are not lower than the maximum price set in accordance with Article 54(1) of Regulation 2015/1222.*"

"*In particular, the implementation of a system of administrative scarcity pricing as described in recital (16)(e) **ensures that prices will be high at times of scarcity and enhances the confidence of future capacity providers that their availability at times of scarcity will be duly rewarded**.*"

"*The Commission considers that, by removing any bidding restrictions **and introducing an administrative scarcity pricing function** before the first delivery* *year of the capacity mechanism, the Polish authorities are* ***acting to reduce the missing money problem*** *and reinforce both the short term availability and long term investment signals sent by the energy market.*"

5.2 Affected Products

The proposal that we develop in this project for the Swedish system targets the introduction of a single ORDC for a single reserve product. We specifically focus on mFRR in our analysis. The introduction of scarcity pricing to multiple reserve products introduces a level of complexity that we discuss here, but without aiming at addressing the challenges that emerge. We focus on two elements: (i) one-way substitutability, and (ii) the fact that certain reserve products are activated by automatic controllers. The two elements interact, as we explain in detail later in this section.

One-way substitutability is a term in economics which refers to the fact that a certain high-quality product can be used for covering the demand of the market for low as well as high-quality products, but not the other way around (i.e. low-quality products cannot be used for covering the demand of the market for high-quality products). This becomes relevant in the context of reserves because quality is measured by the response speed of certain technologies. For instance, hydro units are technologically capable of covering the needs of the system for both aFRR as well as mFRR. On the other hand, certain technologies do not qualify for offering aFRR services, but they are sufficiently fast to cover the needs of the system for mFRR. Note that batteries and some specific demand side management resources may be an exception to this definition. Batteries can arguably deliver aFRR, however their storage limit may obstruct the provision of mFRR services. We point out that US markets do not apply scarcity pricing to Automatic Generation Control (the closest analogue to aFRR). Thus, when considering multiple products, it may be more meaningful to consider mFRR and Replacement Reserve (RR). Nevertheless, we retain the notation aFRR in this section for the sake of illustration. Ultimately, what matters is relations of substitutability between reserve products, because these affect pricing in a non-trivial way, and this is the focus of the present section.

We consider the following market model with independent valuations[14], with the notation being summarized in Appendix section D:

$$max_{p,d,d^{aFRR},d^{mFRR},r^{aFRR},r^{mFRR}} V \cdot d - \sum_{g \in G} C_g \cdot p_g + \int_{x=0}^{d^{aFRR}} V^{aFRR}(x)dx + \sum_{z \in Z} \int_{x=0}^{d^{mFRR}} V^{mFRR}(x)dx \quad (5.2.1)$$

$$(\mu_g): p_g + r_g^{aFRR} + r_g^{mFRR} \leq P_g, g \in G \quad (5.2.2)$$

$$(\nu): d \leq D \quad (5.2.3)$$

[14] The following quantitative analysis is based on [33].

$$(\lambda): d - \sum_{g \in G} p_g = 0 \qquad (5.2.4)$$

$$(\lambda^{aFRR}): d^{aFRR} - \sum_{g \in G} r_g^{aFRR} = 0 \qquad (5.2.5)$$

$$(\lambda^{mFRR}): d^{mFRR} - \sum_{g \in G} r_g^{aFRR} - \sum_{g \in G} r_g^{mFRR} = 0 \qquad (5.2.6)$$

$$p_g \geq 0, d \geq 0, r_g^{aFRR} \geq 0, r_g^{mFRR} \geq 0, d^{aFRR} \geq 0, d^{mFRR} \geq 0 \qquad (5.2.7)$$

The objective function of equation (5.2.1) now considers two ORDCs, one for aFRR and one for mFRR. Note that we consider the case of *independent valuations* for aFRR and mFRR capacity, i.e. the valuation of the TSO for these two products can be expressed in a separable way (in other words, the valuation for a bundle of aFRR and mFRR capacity is the sum of a valuation for aFRR and a valuation for mFRR). This is justified by the fact that the EU balancing market operates two separate platforms for balancing energy in real time (MARI and PICASSO). Note, also, that we consider a single valuation for energy, V. This simplifying assumption is aimed at keeping the analysis tractable.

Constraint (5.2.6) indicates the one-way substitutability of the model: aFRR capacity can contribute towards covering aFRR demand, but it can also contribute towards covering mFRR demand (as indicated in the red term of the constraint).

The KKT system can be described as follows (equalities are not repeated):

$$0 \leq \mu_g \perp P_g - p_g - r_g^{aFRR} - r_g^{mFRR} \geq 0, g \in G \qquad (5.2.8)$$

$$0 \leq \nu \perp D - d \geq 0 \qquad (5.2.9)$$

$$0 \leq p_g \perp C_g + \mu_g - \lambda \geq 0, g \in G \qquad (5.2.10)$$

$$0 \leq r_g^{aFRR} \perp \mu_g - \lambda^{aFRR} - \lambda^{mFRR} \geq 0, g \in G \qquad (5.2.11)$$

$$0 \leq r_g^{mFRR} \perp \mu_g - \lambda^{mFRR} \geq 0, g \in G \qquad (5.2.12)$$

$$0 \leq d^{aFRR} \perp \lambda^{aFRR} - V^{aFRR}(d^{aFRR}) \geq 0 \qquad (5.2.13)$$

$$0 \leq d^{mFRR} \perp \lambda^{mFRR} - V^{mFRR}(d^{mFRR}) \geq 0 \qquad (5.2.14)$$

We assume that the ORDCs for aFRR and mFRR are such that $d^{aFRR} > 0$ and $d^{aFRR} > 0$, meaning that the ORDCs for aFRR and mFRR exceed, at a certain point, the valuation V for energy. The goal of this assumption is to ensure that the price for aFRR and mFRR is essentially implied by the demand curves of these reserve products, which follows straightforwardly from conditions (5.2.13) and (5.2.14).

In order to simplify the analysis and convey the main concepts, let us consider a single generator. We consider the following cases, in order of decreasing severity:

- Case 1: There is a shortage in energy, i.e. $d < D$. In this case, $\nu = 0$. And since $d > 0$, we have $\lambda = V$. Moreover, since $p > 0$ we have $\mu = \lambda - C$. And since $r^{aFRR} > 0$, we conclude that $\mu = \lambda^{aFRR} + \lambda^{mFRR}$. Finally, if it were the case that $r^{mFRR} > 0$ we would have that $\mu = \lambda^{mFRR}$, which would require that $\lambda^{aFRR} = 0$. This, however, is not possible, because we know that $\lambda^{aFRR} = V^{aFRR}(d^{aFRR}) > 0$. So we can conclude that $r^{mFRR} = 0$. What we are interested is the implication on prices. Concretely, we have that the aFRR and mFRR prices are determined by the ORDCs at the level of reserve that is served by aFRR supply, namely $\lambda^{aFRR} = V^{aFRR}(r^{aFRR})$, $\lambda^{mFRR} = V^{mFRR}(r^{aFRR})$. And the energy price is given by $\lambda = \lambda^{aFRR} + \lambda^{mFRR} + C$ which is also equal to V, i.e. $\lambda = V$.
- Case 2: There is no shortage in energy (i.e. $d = D$), but there is no excess of mFRR (i.e. $r^{mFRR} = 0$). In this case, $p > 0$ implies $\mu = \lambda - C$. And also $r^{aFRR} > 0$ implies that $\mu = \lambda^{aFRR} + \lambda^{mFRR}$. Using the same reasoning as in case 1, we can argue that $r^{mFRR} = 0$. Thus, the reserve prices are computed as in case 1. What changes in case 2 relative to case 1 is that the energy price is no longer set by the VOLL, but rather by the following identity: $\lambda = \lambda^{aFRR} + \lambda^{mFRR} + C$.

We revisit each of these cases in a simple example with a generator that has a capacity of $P = 100\ MW$ and a marginal cost of $C = 20\ €/MWh$. The system has a demand D, with a valuation of $V = 1000\ €/MWh$. The aFRR ORDC has the functional form $V^{aFRR}(d^{aFRR}) = c^{aFRR}/d^{aFRR}$. We calibrate this demand curve by requiring it to amount to 500 €/MWh at 40 MW, which implies that $c^{aFRR} = 20000$. We similarly calibrate an mFRR demand curve with a valuation of $V^{mFRR}(d^{mFRR}) = c^{mFRR}/d^{aFRR}$ so that it amounts to a valuation of 200 €/MWh at 40 MW, which implies that $c^{mFRR} = 8000$. We now consider each of the cases analyzed above.

- Case 1: For $D = 80\ MW$, we have that $d = 73.2\ MW$. mFRR supply is indeed 0 MW. aFRR supply is 26.8 MW, which is indeed the amount of demand that is covered for both aFRR and mFRR. The energy price is 1000 €/MWh. But this also coincides with the price of aFRR (700 €/MWh) plus the price of mFRR (280 €/MWh) plus the marginal cost of the marginal unit (20 €/MWh).
- Case 2: For $D = 70\ MW$, we have that $d = 70\ MW$. mFRR supply is indeed 0 MW. aFRR supply is 30 MW, which is indeed the amount of demand that is covered for both aFRR and mFRR. The energy price is 953.3 €/MWh, which is the marginal cost of the marginal unit (20 €/MWh) plus the aFRR adder (666.7 €/MWh) plus the mFRR adder (266.7 €/MWh).

The overall insights of our analysis are presented in Table 11.

	λ	λ^{aFRR}	λ^{mFRR}
Case 1: load shedding	$V = \lambda^{aFRR} + \lambda^{mFRR} + C$	$V^{aFRR}(r^{aFRR})$	$V^{mFRR}(r^{aFRR})$
Case 2: no load shedding	$\lambda^{aFRR} + \lambda^{mFRR} + C$	$V^{aFRR}(r^{aFRR})$	$V^{mFRR}(r^{aFRR})$

Table 11: Energy and reserve prices for each level of stress when multiple reserve products cleared.

A notable challenge with these formulas is that the actual level of aFRR is not driven by a dispatch based on co-optimization, but on an automatic controller. This can be problematic, because it can result in seemingly incoherent results. Consider, for instance, what would happen if aFRR is fully depleted due to activation by an automatic controller. This is typically the case, for instance, in Belgium, where aFRR is often saturated. In this case we would have $r^{aFRR} = 0$, which would imply a very large adder for aFRR, while abundant mFRR capacity may be available. This is clearly incoherent, and stems from the fact that, in actual operations, the dispatch of units is not based on co-optimization.

We are thus tasked with proposing a procedure which approximates as closely as possible the outcome of a co-optimization, while aiming to also approximate the business rules implied by the co-optimization. In order to achieve this objective, we can apply the following rules of thumb, using the output of an energy-only balancing platform:

- For case 1, we set the energy price to VOLL. The mFRR adder is determined by the value of the ORDC for mFRR at the level of leftover reserve in the system. The price of aFRR is set at VOLL minus the price of mFRR minus the price of the energy-only balancing platform (i.e. the marginal cost of the marginal unit).
- An alternative for case 1 would be to set the price of aFRR at the aFRR ORDC evaluated at the total level of reserve in the system.
- For case 2, we first compute the mFRR adder as the ORDC computed at the total level of reserve in the system. We can then compute the aFRR adder by evaluating the aFRR ORDC at the total level of reserve in the system. The energy price is then set at the marginal cost of the marginal unit plus the aFRR adder plus the mFRR adder.

We note that the analysis above over-simplifies the problem, because we do not consider resources that are sufficiently fast to offer mFRR but are not fast enough to offer aFRR. Thus, the model always chooses $r^{mFRR} = 0$. The goal of this development is not to resolve this issue, but rather highlight the challenges that emerge when considering multiple reserve products.

We finally comment on how load shedding is implemented in practice and what it implies in terms of pricing, given existing or future arrangements. When load is shed in the Nordics, there are rules for distributing load shedding between areas. MARI and PICASSO have different mechanisms foreseen for handling shortage. The current practice in the Nordics is summarized in 1.3. In summary, when the system runs out of capacity the price is set to the price of the last selected bid. Price setting in the case of mobilization of strategic reserve is further detailed in section 1.3.

5.3 Compatibility with CRMs

Scarcity pricing can coexist with capacity remuneration mechanisms. Figure 32 presents a mapping of ORDCs and capacity auctions[15] in European Member States. One can observe in the figure that the two mechanisms can (and in certain cases already do) coexist. Regarding the CRMs that are presented in Figure 32, the mapping does not include capacity payments (e.g. Spain) or strategic reserve.

Figure 32: Mapping of ORDCs and CRMs in European Member States. A green box indicates that the mechanism in question has already been implemented in the concerned Member State, whereas an orange box indicates that the mechanism is under consideration.

The coexistence of scarcity pricing and CRMs does not imply double payment in a perfectly competitive market. We establish this argument formally in appendix C of the present report. The intuition of the argument (which also applies to scarcity pricing) is that, as long as the energy market does not result in missing money, the capacity market can indeed be in place, but its clearing price tends to zero since there is no missing money to be recovered by the mechanism. In order for this result to hold, it is important that the CRM demand curve should not be oversized. We state the formal result below (and prove it in the appendix), and we then proceed to discuss the shape of CRM demand curves and how the CRM interacts with the energy market.

The formal result that is proven in the appendix is the following:

Proposition 5.3.1: If the capacity auction demand curve has a price of 0 €/MWh at the optimal level of capacity for the ORDC design, then the ORDC optimal capacity mix is also optimal for the ORDC+CRM design, and the capacity price is 0 €/MWh.

Note that CRM auction demand curves follow a typical shape along the lines of Figure 33. The typical demand curve thus consists of a part with a high valuation, which is specifically a multiple of the cost of new entry. This horizontal part typically extends to 100% of the target capacity of the system, as determined by adequacy analyses. The demand curve then slopes of linearly to 0 €/MWh at a multiple of the target capacity of the system. For instance, in the left panel of Figure 33 we observe that the demand curve evaluates to 0 €/MWh at 115% of the target capacity.

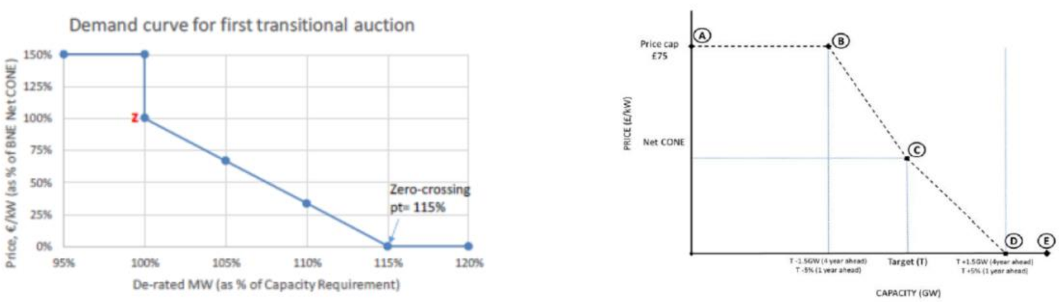

Figure 33: Irish (left) and UK (right) CRM demand curves. Source: [31], [32].

[15] Note that there exist capacity remuneration mechanisms beyond capacity auctions in the EU market, which aim to remunerate capacity and ensure adequacy, including capacity payments and strategic reserves.

The motivation of the left horizontal part of the CRM demand curve is to ensure that the system procures at least the target capacity, as determined by an adequacy study. The motivation of the linear sloped part is to avoid a bimodal behavior which was observed in early capacity auctions in the US where inelastic (i.e. vertical) CRM demand curves were used. In these CRM auctions, the clearing price would behave in a binary way: if the supply were below the target level of the auction the clearing price would be equal to a multiple of cost of new entry (CONE), while if the offered capacity exceeded the target capacity of the system then the CRM price would collapse to 0 €/MWh.

Note that proposition 5.3.1 challenges the shape of the CRM demand curve in the left panel of Figure 33. Concretely, the fact that the CRM demand curve extends beyond the target capacity of the system in Figure 33 raises a risk of over-dimensioning the system. Instead, the proposition motivates a CRM demand curve in which the valuation of the demand curve becomes 0 €/MWh at the level of target capacity, and not beyond it.

In its review of the Polish CRM [38], the European Commission exactly reflects the intuition of proposition 5.3.1. Concretely, one reads in recital 156:

"*The Commission considers that, by removing any bidding restrictions* ***and introducing an administrative scarcity pricing function*** *before the first delivery year of the capacity mechanism, the Polish authorities are* ***acting to reduce the missing money problem*** *and reinforce both the short term availability and long term investment signals sent by the energy market.*"

6. Cross-Border Effects and Pan-European Balancing Platforms

Since scarcity adders are expected to interact with the settlement prices produced by MARI, it is clear that these adders cannot rely on co-optimization, as MARI is an energy-only platform. There is no co-optimization of energy and reserves foreseen in MARI, nevertheless the ERCOT experience demonstrates that co-optimization is not a necessary condition for the implementation of scarcity pricing.
In section 3.4 we provide a proof of concept for how Sweden can integrate its multi-zone operation with the MARI platform. In this section we rather focus on the incentives of market participants as a consequence of introducing scarcity adders in Sweden. Thus, we highlight certain elements of what the introduction of adders in order to approximate the outcome of a co-optimization implies in terms of cross-border energy-only platforms.

Before advancing to the detailed discussion of the incentives implied by the different design proposals, it is firstly important to remind what is meant by introducing adders. In its public consultation [21], ELIA considers two possible approaches for integrating scarcity prices in cross-border platforms: (i) applying adders ex-post to the MARI prices, or (ii) computing adders based on anticipated levels of reserves, and uplifting the offers of BSPs to the MARI platform before these offers are submitted to the platforms. The second approach has never been proposed or analyzed by Papavasiliou [5, 12], and it is not clear whether it is institutionally viable or compatible with first principles. The discussion in this section is rather focused on the first option, which is analyzed extensively in section 4 as well as the CREG proposal [6] for the implementation of scarcity pricing. The point is illustrated in Figure 34.

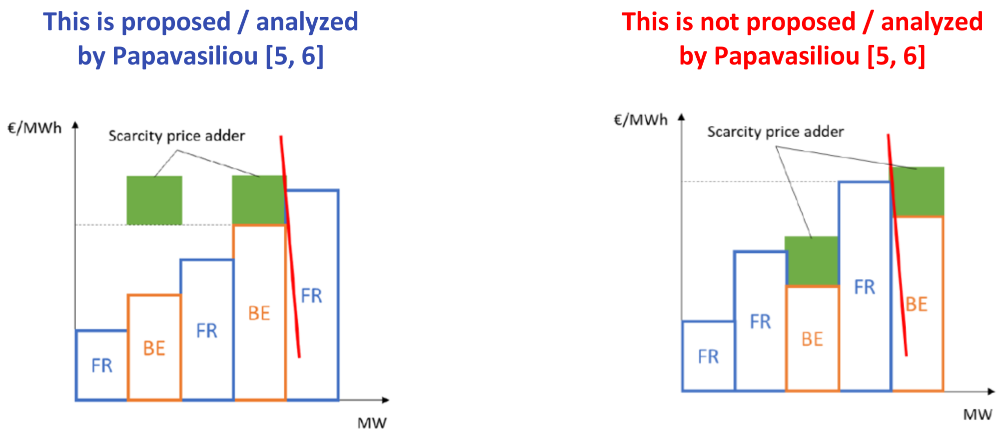

Figure 34: ELIA [21] describes two interpretations of the proposal of Papavasiliou [5, 12]. The first interpretation is that adders apply ex post to platform prices (left panel). The second interpretation is that adder apply ex ante to the bids of the Member State which implements scarcity pricing (right panel). The correct interpretation is the first one.

6.1 Energy-Only Platforms and Co-Optimization Can Produce Different Outcomes [23]

Before advancing to a discussion of bidding incentives and financial transfers, we first underline the fact that energy-only platforms and scarcity pricing can produce different dispatch outcomes. In order to illustrate this point, we use the example in Figure 35. In this example, we have the BE zone implementing scarcity pricing. The Belgian BSP submits a linked bid for energy and reserve at a total capacity of 10 MWh, and with a marginal energy cost of 10 €/MWh. The TSO in BE procures reserve up to 20 MWh at 100 €/MWh. The FR zone has one BSP, which has a capacity of 10 MWh and a marginal cost of 20 €/MWh. Zone BE is short by 8 MWh. The transmission capacity from FR to BE is 3 MW.

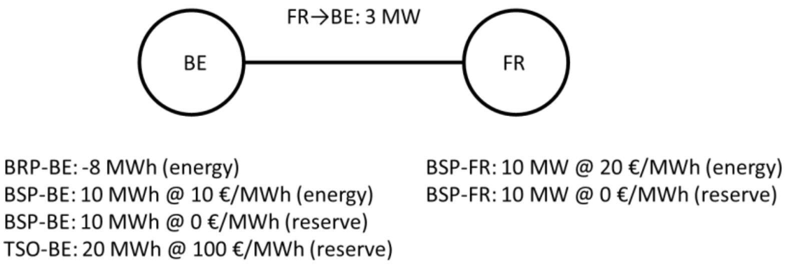

Figure 35: Two-zone example used to illustrate that an energy-only platform and co-optimization models can produce different dispatch outcomes.

In an energy-only market, BSP-BE is activated fully at 8 MWh so that it can entirely 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 fully satisfy BRP-BE are imported from BSP-FR.

Note that the dispatch is different between the two models. The energy supply of BSP-BE is held back in co-optimization, so that the capacity can be used for covering 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 price of reserve 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 would implement scarcity pricing unilaterally, it would seek to price energy and reserve so as 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 reserves are anyway covered by the most expensive units, therefore one requires more complex interactions (e.g. related to maximum reserve limits of individual units) in order to disrupt merit-order dispatch [39].

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 anyway guaranteed to respect their no-arbitrage conditions by design of the balancing platform. Concretely, BSP-FR does not require a scarcity adder in order to deliver energy to the platform: the energy price of the platform for FR is already sufficient for this purpose.

6.2 Bidding Incentives and Efficiency of Dispatch [6]

The examples of this section are based on section 5.5 of the CREG scarcity pricing note [6]. The prototypical example that we consider is presented in Figure 36. In this example, we consider a two-zone system. Zone A does not apply scarcity pricing, whereas zone B applies scarcity pricing. There is an imbalance of -50 MW in zone A, and -700 MW in zone B. Zone A has a BSP that has a generation capacity of 1000 MW at a (privately known) marginal cost of 100 €/MWh. There is also a BSP in zone B with a capacity of 1000 MW at a (privately known) marginal cost of 200 €/MWh. We analyze how the BSPs choose to bid, when behaving competitively, under the different designs that we have discussed in section 4.

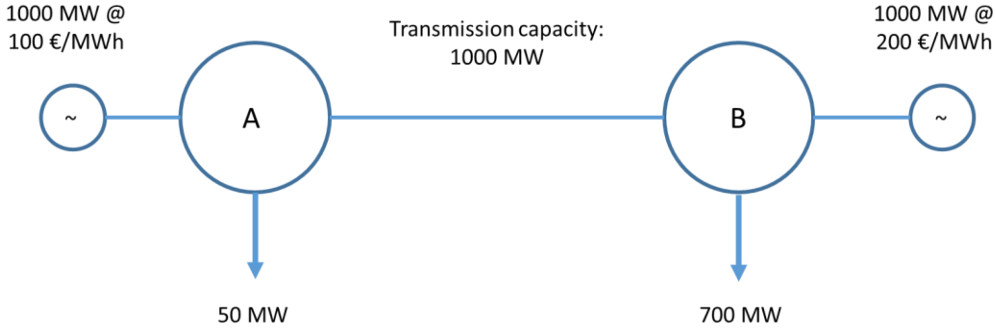

Figure 36: Cross-border example considered for illustrating bidding incentives.

6.2.1 No adder

In this case, the optimal strategy of agents is for them to bid their entire capacity at their true marginal cost. The resulting system is dispatched by having the generator in zone A covering the entire imbalance of the system, i.e. generator A produces 750 MWh. The dispatch is optimal, and efficiency is not undermined.

6.2.2 Applying Adders on Imbalances Only

In this case, the optimal strategy for an agent in the B zone is to self-dispatch so as to earn the adder which is only applicable to imbalance energy. Concretely, let us assume that the adder is 200 €/MWh in zone B because zone B is tight, and so long as zone B is long. Then, it is in the interest of BSPB to self-dispatch long (for instance up to 700 MWh, i.e. up to the point where it neutralizes the imbalance of zone B). This is clearly inefficient, and thus contravenes article 3(m) of the CEP [22].

6.2.3 A Disciplined Approximation of Co-Optimization

In this case, as in the case of section 6.2.1, the optimal strategy of BSPs is for them to bid their entire capacity at its true marginal cost. Let us further assume that zone B is tight in this example. In this case, the generator in zone A produces 750 MWh, and the optimal dispatch is attained. Note that the generator in zone B does not have an incentive to underbid, because it is indifferent between receiving the adder for energy or for its spare reserve capacity. Thus, this design does not distort optimal dispatch. The balancing and imbalance price in zone A becomes 100 €/MWh, while it becomes 300 €/MWh in zone B after the adder is applied.

6.3 TSO Cash Flows [6]

We now proceed to discuss the financial exposure of the TSO in the case of cross-border activation. The examples discussed in this section are sourced from CREG [6] and ELIA [21]. The examples are illustrated graphically in Figure 38.

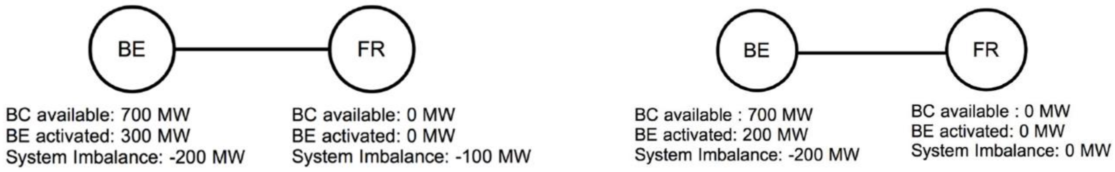

Figure 38: Illustrative examples used for analyzing the financial exposure of TSOs in the case of cross-border activations. In the left panel, the BE system is exporting balancing energy to FR. In the right panel, the BE system is not exporting any balancing energy to FR.

We assume throughout these examples that BE implements scarcity pricing unilaterally, following the design of section 4.4. In the left panel of Figure 38, both BE and FR are short. The BSP in BE, which has a total capacity of 700 MW, is activated in order to cover 200 MW of imbalance in BE and 100 MW of imbalance in FR. In the right panel of Figure 38, only BE is short. The BSP in BE is now activated in order to only cover 200 MW of imbalance in BE.

The discussion focuses on the financial exposure of the TSO in the zone which implements scarcity adders, namely BE. The resulting cash flows are illustrated in Table 12. Note that, in both cases, the TSO is liable for $-\lambda^R \cdot 500$. Concretely, in both cases, TSO-BE pays for the difference between its local reserve capacity and the local imbalance at the adder price:

- In the case of example 1, this amount corresponds to the payment for the excess of reserve capacity in BE, as well as the energy activated for covering the imbalance in FR.
- In the case of example 2, this amount only corresponds to the payment for the excess of reserve capacity in BE.

	BRP FR	BRP BE	BSP activated	BSP non-activated	TSO
Example 1	$-\lambda^B \cdot 100$	$-(\lambda^B + \lambda^R) \cdot 200$	$(\lambda^B + \lambda^R) \cdot 300$	$\lambda^R \cdot 400$	$-\lambda^R \cdot 500$
Example 2	0	$-(\lambda^B + \lambda^R) \cdot 200$	$(\lambda^B + \lambda^R) \cdot 200$	$\lambda^R \cdot 500$	$-\lambda^R \cdot 500$

Table 12: Settlements in the two examples of section 6.3.

References

[1] Hogan, W. (2013). “Electricity scarcity pricing through operating reserves.” *Economics of Energy and Environmental Policy,* vol. 2, no. 2, pp. 65–86. https://doi.org/10.5547/2160-5890.2.2.4

[2] A. Papavasiliou, Y. Mou, “Modeling Energy-Only Markets in the Presence of Shortage Pricing and Capacity Remuneration Mechanisms”, working paper, 2022

[3] A. Papavasiliou, Y. Smeers, “Remuneration of Flexibility under Conditions of Scarcity: A Case Study of Belgium”, the Energy Journal, vol. 38, no. 6, pp. 105-135, 2017

[4] ERCOT, 2014. ERCOT Market Training: Purpose of ORDC, Methodology for Implementing ORDC, Settlement Impacts for ORDC, TX: ERCOT.

[5] A. Papavasiliou, Y. Smeers, G. de Maere d'Aertrycke, Study on the general design of a mechanism for the remuneration of reserves in scarcity situations, June 6, 2019

[6] Belgian Commission for the Regulation of Electricity and Gas (CREG), “Study on the implementation of a scarcity pricing mechanism in Belgium”, October 2022, available online: https://www.creg.be/sites/default/files/assets/Publications/Studies/F2203EN.pdf

[7] J. Cartuyvels, A. Papavasiliou, “A market simulation methodology for the calibration of ORDC”, technical report prepared on behalf of the Belgian regulatory commission for electricity and gas (CREG), June 2021

[8] W. Hogan, “Scarcity Pricing and Locational Operating Reserve Demand Curves”, https://scholar.harvard.edu/whogan/files/hogan_ferc_060210.pdf

[9] J. Cartuyvels, A. Papavasiliou, Calibration of Operating Reserve Demand Curves Using a System Operation Simulator, forthcoming in the IEEE Transactions on Power Systems
https://ap-rg.eu/wp-content/uploads/2022/08/J40.pdf

[10] Svenska kraftnat, “Kraftbalansen på den svenska elmarknaden, rapport 2021”, May 28, 2021

[11] NSIDE, “Co-Optimization of Energy and Balancing Capacity in the European Single Day-Ahead Coupling Roadmap study”, May 18, 2022

[12] A. Papavasiliou, “Scarcity Pricing and the Missing European Market for Real-Time Reserve Capacity”, the Electricity Journal, vol. 33, no. 10, September 2020.

[13] A. Papavasiliou, Y. Smeers, G. de Maere d'Aertrycke, “Market Design Considerations for Scarcity Pricing: A Stochastic Equilibrium Framework”, The Energy Journal, vol. 42, no. 5, pp. 195-220, 2021.

[14] Cramton, Peter C., and Steven Stoft. "The convergence of market designs for adequate generating capacity with special attention to the CAISO's resource adequacy problem." (2006).

[15] Stoft, S., 2002. Power system economics: designing markets for electricity. Piscataway: IEEE press.

[16] R. S. Sutton and A. G. Barto, Reinforcement learning: an introduction. MIT press, 2018.

[17] A. Papavasiliou, G. Bertrand, “Market Design Options for Scarcity Pricing in European Balancing Markets”, IEEE Transactions on Power Systems, vol. 36, no. 5, pp. 4410-4419, September 2021.

[18] A. Papavasiliou, “Analytical Derivation of Optimal BSP / BRP Balancing Market Strategies”, https://perso.uclouvain.be/anthony.papavasiliou/public_html/AnalyticalV3.pdf

[19] G. Bertrand, “Enhancing the Value of Flexible Capacity through Continuous Intraday Markets and Scarcity Pricing”, PhD thesis, UCLouvain

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

[21] ELIA, 2020. Final report on Elia’s findings regarding the design of a scarcity pricing mechanism for implementation in Belgium, Brussels: ELIA.

[22] 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.

[23] A. Papavasiliou , G. Bertrand, A. Marien, J. Cartuyvels, “Implementation of Scarcity Pricing without Co-Optimization in European Energy-Only Balancing Markets”, under review

[24] Boiteux, M., 1960. Peak-load pricing. The Journal of Business, 33(2), pp. 157-179.

[25] Fabra, N., 2018. A primer on capacity mechanisms. Energy Economics, Volume 75, pp. 323-335.

[26] Hogan, W., 2005. On an ‘energy only’ electricity market design for resource adequacy, Cambridge, MA: Center for Business and Government, JFK School of Government, Harvard University.

[27] A. Papavasiliou, CREG Scarcity Pricing Workshop, Principles of Scarcity Pricing, Brussels, Belgium, April 29, 2022

[28] European Commission, 2017. Commission regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing, Brussels: European Commission.

[29] N-SIDE, AFRY, “Cross-Zonal Capacity Allocation with Co-Optimization”, November 2020

[30] ACER, 2020. Annex I to 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., Ljubljana: ACER.

[31] European Commission. State aid no. SA. 44464 (2017/N) - Ireland: Irish capacity mechanism. Technical report, 2017.

[32] European Commission. State aid SA. 35980 (2018/NN) - United Kingdom electricity market reform: Capacity mechanism. Technical report, 2019.

[33] A. Papavasiliou, “Implementation of Scarcity Pricing in EU Balancing Markets”, technical report, March 5, 2021

[34] Bemporad, Alberto, Carlo Filippi, and Fabio D. Torrisi. “Inner and outer approximations of polytopes using boxes.” Computational Geometry 27.2 (2004): 151-178

[35] Federal Public Service Economy, 2019. Belgian electricity market: implementation plan, Brussels: Federal Public Service Economy.

[36] 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, Brussels: European Commission.

[37] European Commission, 2018. State Aid SA.42011 (2017/N) – Italy – Italian Capacity Mechanism, Brussels. Available online: https://ec.europa.eu/competition/elojade/isef/case_details.cfm?proc_code=3_SA_42011.

[38] European Commission, 2018. *State aid No. SA.46100 (2017/N) – Poland – Planned Polish capacity mechanism,* Brussels: Decision Public Version.

[39] NYISO, 2019. Ancillary services shortage pricing, NYISO: New York ISO.

[40] A. Papavasiliou, Y. Smeers, G. Bertrand, An Extended Analysis on the Remuneration of Capacity under Scarcity Conditions, Economics of Energy and Environmental Policy, vol. 7, no. 2, 2018. Available online: https://ap-rg.eu/wp-content/uploads/2020/07/J15.pdf

[41] Belgian regulatory commission for electricity and gas (CREG), Décision relative à la proposition d’ELIA TRANSMISSION BELGIUM portant modification des règles d’équilibrage pour la compensation des déséquilibres quart-horaires, July 19, 2022.

[42] Tennet, “Proposal for updated pricing policy for balancing energy and imbalance”, Consultation from 22-08-2022 to 14-10-2022

[43] ELIA, 2018. Study report on Scarcity Pricing in the context of the 2018 discretionary incentives, Brussels: ELIA.

[44] Zarnikau, Jay, S. Zhu, Chi Keung Woo, and C. H. Tsai. "Texas's operating reserve demand curve's generation investment incentive." Energy Policy 137 (2020): 111143.

[45] A. Papavasiliou, Modeling Cross-Border Interactions of EU Balancing Markets: a Focus on Scarcity Pricing, chapter in Mathematical Modelling of Contemporary Electricity Markets, Elsevier, edited by Athanasios Dagoumas, ISBN: 9780128218389 https://ap-rg.eu/wp-content/uploads/2021/01/B3.pdf

Appendix A: Notation of Model (1.2)

In this section we present the notation that is used in the two-stage stochastic dispatch model that is employed in section 1 for deriving the ORDC adder formulas that are based on VOLL and LOLP.

Sets
Ω: set of scenarios
G: set of flexible production units in the system

Variables
d: demand satisfied in the first stage
p_g: production of flexible unit g in the first stage
r_g: reserve capacity of flexible unit g in the first stage
δ_ω: amount of imbalance demand that is satisfied in scenario ω

Parameters and functions
$MB(\cdot)$: marginal benefit function
$MC_g(\cdot)$: marginal cost function of unit g
$VOLL$: value of lost load
$\widehat{MC}(\cdot)$: approximation of system incremental cost for meeting an additional increment of demand
P_g: nominal rating of unit g
Δ_ω: imbalance under scenario ω
Π_ω: probability of scenario ω

Appendix B: Notation of Model (3.5)

In this section we present the notation that is used in the energy / reserve / transmission co-optimization model that is proposed in section 3.4 so as to introduce an ORDC to a multi-zone model of the Swedish system.

Sets
G: set of flexible production units in the system
Z: set of zones in the system
G_z: set of generators in zone z
K: set of links

Variables
d_z: demand served in zone z
p_g: production of flexible unit g
r_g: reserve capacity of flexible unit g
dr: amount of reserve demand that is satisfied
f_k: flow on link k
$fR_k^{+/-}$: reserve flow in the reference direction / opposite to the reference direction of link k

Parameters and functions
$VOLL$: value of lost load
MC_g: marginal cost function of unit g
$ORDC(\cdot)$: operating reserve demand curve
P_g: nominal rating of unit g
R_g: reserve capacity of unit g
D_z: inelastic demand in zone z

Appendix C: Proof of Coexistence between Scarcity Pricing and CRMs [2]

In this section we prove the following proposition, which is already stated in section 5.3.

Proposition 5.3.1: If the capacity auction demand curve has a price of 0 €/MWh at the optimal level of capacity for the ORDC design, then the ORDC optimal capacity mix is also optimal for the ORDC+CRM design, and the capacity price is 0 €/MWh.

Proof

In order to prove our result, we consider the following capacity expansion model.

$$\max_{x,p,d,xd,r,dR} \sum_{t\in T} \Delta T_t \cdot (V \cdot d_t - \sum_{i\in I} MC_i \cdot p_{it}) + \sum_{l\in LR, t\in T} VR_{lt} \cdot \Delta T_t \cdot dR_{lt} + \sum_{l\in LC} VC_l \cdot xd_l - \sum_{i\in I} IC_i \cdot x_i \tag{C.1}$$

$$\sum_{l\in LC} xd_l - \sum_{i\in I} x_i = 0 \qquad (\lambda C) \tag{C.2}$$

$$d_t - \sum_{i\in I} p_{it} = 0, t \in T \qquad (\Delta T_t \cdot \lambda_t) \tag{C.3}$$

$$\sum_{l\in LR} dR_{lt} - \sum_{i\in I} r_{it} = 0, t \in T \qquad (\Delta T_t \cdot \lambda R_t) \tag{C.4}$$

$$p_{it} + r_{it} \leq x_i, i \in I, t \in T \qquad (\Delta T_t \cdot \mu_{it}) \tag{C.5}$$

$$d_t \leq D_t, t \in T \qquad (\Delta T_t \cdot \nu R_t) \tag{C.6}$$

$$xd_l \leq DC_l, l \in LC \qquad (\nu_l) \tag{C.7}$$

$$dR_{lt} \leq DR_{lt}, l \in LR, t \in T \qquad (\Delta T_t \cdot \nu_{lt}) \tag{C.8}$$

$$x, p, d, xd, dR, r \geq 0 \tag{C.9}$$

In this capacity expansion model, green terms correspond to the capacity market and red terms correspond to a scarcity pricing mechanism. This is a standard capacity expansion planning model. The notation of the model is explained in the end of this section.

Take the KKT conditions of the ORDC+CRM model, and verify that they are satisfied for exactly the same capacities as the ORDC model, with the capacity price set to zero.

$$\sum_{l\in LC} xd_l - \sum_{i\in I} x_i = 0 \qquad (\lambda C)$$

$$d_t - \sum_{i\in I} p_{it} = 0, t \in T \qquad (\Delta T_t \cdot \lambda_t)$$

$$\sum_{l\in LR} dR_{lt} - \sum_{i\in I} r_{it} = 0, t \in T \qquad (\Delta T_t \cdot \lambda R_t)$$

$$0 \leq \Delta T_t \cdot \mu_{it} \perp p_{it} + r_{it} \leq x_i, i \in I, t \in T$$

$$0 \leq \Delta T_t \cdot \nu R_t \perp d_t \leq D_t, t \in T$$

$$0 \leq \nu_l \perp xd_l \leq DC_l, l \in LC$$

$$0 \leq \Delta T_t \cdot \nu_{lt} \perp dR_{lt} \leq DR_{lt}, l \in LR, t \in T$$

$$0 \leq x_i \perp IC_i - \lambda C - \sum_{t \in T} \Delta T_t \cdot \mu_{it} \geq 0, i \in I$$

$$0 \leq p_{it} \perp MC_i - \lambda_t + \mu_{it} \geq 0, i \in I, t \in T$$

$$0 \leq d_t \perp -V + \lambda_t + \nu_{lt} \geq 0, t \in T$$

$$0 \leq dR_{lt} \perp -VR_{lt} + \lambda R_t + \nu_{lt} \geq 0, l \in LR, t \in T$$

$$0 \leq r_{it} \perp -\lambda R_t + \mu_{it} \geq 0, i \in I, t \in T$$

$$0 \leq xd_l \perp -VC_l + \nu_l \geq 0, l \in LC$$

The green terms indicate the part of the KKT conditions that is associated with the CRM. If we remove the green terms from the model, we have a set of conditions which characterize the EOM+ORDC optimal solution.
Let us consider the following vector:

- All ORDC primal and dual variables are equal to the primal – dual optimal solution of the ORDC model
- $\lambda C = 0$
- $xd_l = DC_l$ if segment l is in the money, $xd_l = 0$ if the segment l is out of the money, and xd_l such that $\sum_{l \in LC} xd_l - \sum_{i \in I} x_i = 0$ if the segment l is at the money.
- Set the multiplier ν_l equal to the profit margin of segment l.

Note that this vector satisfies the optimality conditions of the ORDC+CRM problem by construction, but it is possible to construct it because the demand curve has a valuation equal to zero at the level of capacity which corresponds to the amount of capacity built under the ORDC model. Thus, the optimal solution of the ORDC model is also optimal for the ORDC+CRM model, and the capacity market has an equilibrium price of 0 €/MWh, so it is as if it does not even exist.

■

The notation that is used in this capacity expansion model is summarized below.

Sets
T: set of periods
I: set of technologies
RL: set of segments in ORDC
CL: set of segments in CRM demand curve
$IF \subseteq I$: set of technologies with final capacity targets

Parameters
ΔT_t: fraction of duration period $t \in T$
V: valuation of consumers for electricity
MC_i: marginal cost of technology $i \in I$
D_t: demand of period $t \in T$
IC_i: investment cost of technology $i \in I$

VR_{lt}: willingness to pay of ORDC segment $l \in LR$ for period $t \in T$
DR_{lt}: demand of ORDC segment $l \in LR$ for period $t \in T$
DC_l: demand of CRM segment $l \in LC$
Z_i: zone in which asset i is located
X_i^0: initial capacity of technology $i \in I^0$
X_i^F: target capacity of technology $i \in I^F$

Variables
d_t: demand of period $t \in T$
x_i: capacity investment of technology $i \in I$
p_{it}: production of technology i for covering demand of period $t \in T$
r_{it}: reserve of technology i for covering demand for reserve of period $t \in T$
dr_{lt}: demand for reserve in period $t \in T$ in ORDC segment $l \in LR$
xd_l: capacity demand of CRM demand curve segment $l \in LC$

Appendix D: Notation of Model (3.5)

The notation of the multi-product model can be summarized as follows.

Sets
G: set of flexible units in the system

Parameters and functions
V: VOLL
C_g: marginal cost of flexible unit $g \in G$
$V^{aFRR}(\cdot)$: operating reserve demand curve for aFRR capacity
$V^{mFRR}(\cdot)$: operating reserve demand curve for mFRR capacity
P_g: nominal capacity of flexible unit $g \in G$
D: system load

Variables
p_g: amount of balancing energy provided by unit $g \in G$
d: amount of energy demand
d^{aFRR}: amount of balancing capacity that can respond within the aFRR full activation time that is actually available in real time
d^{mFRR}: amount of balancing capacity that can respond within the aFRR full activation time that is actually available in real time
r_g^{aFRR}: amount of aFRR balancing capacity made available by unit $g \in G$
r^{mFRR}: amount of mFRR balancing capacity made available by unit $g \in G$