Securing investment for essential goods. How to design the demand side of capacity markets?

Older working paper

Leopold Monjoie[1,2,3], Fabien Roques[1,2]

[1] Université Paris Dauphine, PSL Research University, LEDa, Place du Marechal de Lattre de Tassigny, 75016, Paris, France

[2] Chair European Electricity Markets

[3] leopold.monjoie@dauphine.eu

Abstract. This paper studies the provision of a homogeneous good with time-varying uncertain stochastic demand and capacity-constrained producers. Due to market failures and public interventions, private agents typically under-procured investments. We analyze the design of long-term capacity markets where producers can sell their investment availability to restore efficient investment levels. While their supply-side effect on investment decisions is well known, we focus on the demand-side effects generated by their implementation. We provide a novel sequential analytical framework of the capacity market followed by short-term markets (wholesale and retail) to describe how different market design regimes can affect the equilibrium of the system. First, we develop the model regarding the implementation of a single buyer on the capacity market, which needs to choose the cost allocation regime for the demand side. And we extend our model to study how the realized demand is accounted for in the market design. We demonstrate that the ability of the capacity market to restore the social optimum, or at least to reach a second-best optimum, crucially depends on the different design regimes of the capacity market, as well as on the assumptions of policy interventions and the various market inefficiencies.

Keywords: market design · investment decisions · imperfect competition · regulation.

1 Introduction

For some essential goods with demand varying over time, wholesale markets' private incentives are not sufficient to ensure that producers make enough investments to meet peak demand in advance of the time when the peak demand materializes.[4] In such industries, due to the critical importance of these goods, policymakers tend to intervene and implement price caps or other types of regulation that distort the price signal and undermine investment incentives. Moreover, the availability of the production capacity for these goods can be characterized as public goods during scarcity periods, for instance, during a cold wave with peak electricity demand or a pandemic with peak demand for medicine or medical equipment. In such circumstances, the absence of adequacy between the capacity and the peak demand, combined with the difficulty of implementing efficient rationing, leads to high costs for society.[5]

One solution to restore the optimal level of investment lies in implementing a mandatory capacity market in which producers commit to having capacity available to collectively meet the expected peak demand, as the regulator prescribes. Current implementations of such mechanisms have been the prerogative of the electricity sector under the name capacity remuneration mechanism.[6] However, the COVID pandemic prompted interest in setting up such mechanisms for vaccine supply. For instance, in Ahuja et al. (2021), they study the implementation of a capacity price for the procurement of vaccines and state that *"to accelerate the vaccine delivery timetable, buyers should directly fund manufacturing capacity"*. Ockenfels (2021) proposes a hybrid mechanism that combines a capacity remuneration mechanism with guaranteed prices. On the supply side of those mechanisms, each participating producer makes a price-quantity offer for a capacity. If a producer sells capacity in this capacity market, he receives a capacity price and commits to being available to produce over future periods.

[4]Our framework fits into the more general analysis of industries in which a form of competition follows long-run investments, such as electricity markets (De Frutos and Fabra, 2011), communication network (Acemoglu et al., 2009), or radio spectrum (Yan, 2020).

[5]The COVID-19 crisis offers a recent example of systemic cost induced by the lack of productive capacity. The subject is well known in the electricity sector while remaining current, as illustrated by blackouts experienced in China and Texas. Congestion in transport infrastructure can also be directly linked to the discussion in this paper.

[6]See for instance Doorman et al. (2016) for a technical description of potential implementations.

While the supply emerges naturally in those markets, the capacity demand requires a regulatory intervention. Indeed, the public-good nature of investment during high-demand periods implies that consumers are unwilling to buy capacities in capacity markets.[7] Hence, the regulator must define the demand function administratively so the market clears and provides producers' capacity prices. For instance, this demand side of capacity markets can be characterized by a single buyer procuring all the capacity or by a set of rules describing how consumers or retailers participate in the capacity market. This paper establishes a framework describing the economic impacts of different demand side designs for capacity markets and their policy implications. We focus on two interrelated questions that relate to (i) the cost allocation regime, that is, how a single buyer allocates the capacity price between capacity buyers and final consumers, and (ii) the degree to which the final consumers' realized demand is accounted in the market allocation design.[8] In this paper, we describe the channel through which each possible market design impacts the system equilibrium. We show that specific market design can affect the demand side of short-term markets, which also has some feedback effects on capacity market equilibrium and constrains the regulator in choosing the efficient investment level. In other words, our model allows for the analysis of the endogeneity between the welfare-maximizing market outcome regulators aim to restore through their intervention and the design of the implemented capacity market.

The direct effect of an additional stream of remuneration on investment decisions is well understood. The current literature covers a significant range of issues : (i) the outcomes in investment decisions with and without capacity markets, (ii) the effect of market power on the capacity price determination, (iii) the relation with risk and business cycle (iv) the discrimination between different investment technologies.[9] However, to our knowledge, there has been no formal analysis

[7]Transaction cost and asymmetric information prevent adequate transactions up to the optimal level; see, for instance, Keppler et al. (2021) for a discussion in electricity markets. The insurance of having enough capacity has a private value (how much each consumer is willing to pay to avoid inadequacy) and a social value, as an increase in investment reduces the probability of systemic costs (Fabra, 2018). Furthermore, knowing the willingness to pay for this insurance is sometimes technologically, socially, and economically impossible.

[8]We also use interchangeability term ex-post temporality, as capacity markets are set before demand is known; as well as the term capacity demand allocation. At the same time, while this is out of the scope of this paper, the dynamic nature of the provision of an essential good, such as electricity, is central. It includes the decision to invest, which can span from many years to a few months, and the decision to consume the good. For instance, the good is almost immediately consumed for electricity and vaccines. On the other hand, medical equipment, intensive care units, strategic energy reserves, or human capital are more durable goods.

[9]See Bublitz et al. (2019) for a detailed literature review on the theory and implementation issues of capacity markets in electricity markets.

of different capacity markets' demand side designs, the incentive properties of these alternative approaches, and their ability to restore the socially optimal level of investment beyond the direct effect of the increase of the marginal investment value due to the additional capacity price. On the other hand, the importance of the demand function design in the capacity market is well known.[10] However, those papers still only consider the effect of the capacity market directly on the supply side. In contrast, our paper underlines the indirect effect of this instrument on retailers and consumers, which in turn impacts producers. Scouflaire (2019) is the first paper to represent retailers' strategies in the capacity market. She develops a theoretical model to analyze the preferences regarding information precision for uncertain future demand. Contrary to our approach, she models heterogeneous price taker producers and homogeneous buyers competing for *à la Cournot* under uncertainty on their level of capacity obligation. In this paper, we take a step back from the supply-side-focused approach and develop a model that sheds light on the complex interactions between the capacity market design and the demand side. Our model provides some new and non-intuitive insights into their incentive properties and their ability to restore the socially optimal (welfare-maximizing) level of investment.

The central results lie in the relation between the outcome in terms of investment level and expected welfare at the system equilibrium and the choice of a particular demand-side market design. To do so, we extend the canonical benchmark model for a homogeneous good characterized by time-varying demand, which describes the relationship between short-term production and long-term investment decisions.[11] Producers make long-run investments in a single technology in the upstream market to produce a homogeneous good subsequently, given an uncertain future demand. Then, the downstream retailers aggregate and resell the goods at no cost to the final consumers. Our model extends the literature by providing a novel analytical framework that includes a capacity market equilibrium in addition to investment and short-term decisions. We derive an endogenous supply function in the capacity market to do so. Namely, following the main theoretical view for capacity markets[12], we assume that producers offer their marginal opportunity cost of providing

[10]See, for instance, Hobbs et al. (2007) and Bushnell et al. (2017) Fabra (2018) Brown (2018).

[11]This model was first developed in a regulated context by Boiteux (1949) for the electricity sector; it was then extrapolated to a market with private producers by Crew and Kleindorfer (1976). This model is widely used to highlight the risk of underinvestment in production capacity.

[12]See for instance Creti and Fabra (2007).

additional capacity. This opportunity cost equals the marginal loss of revenue incurred by the investment level beyond the profit-maximizing equilibrium. Our modeling proposition is central as any indirect effects generated by the capacity market can affect the expected revenue made by the producers and can indirectly be captured during the formation of the supply function in the capacity market. For the capacity equilibrium to emerge, we make specific assumptions for the demand-side designs. The paper assumes a single buyer that chooses a demand function that maximizes expected social welfare. In an extension, we also characterized the equilibrium when retailers participate in the capacity market.

We start our analysis by introducing a price cap regulation, which can be interpreted as representing the effect of different types of market distortions induced by a range of market failures and regulatory interventions, which are typical for electricity, for instance, and can take the form of price caps or regulations with a similar effect on price dynamics in practice.[13] Such a price cap reduces the expected revenues of producers and undermines investment compared to the level needed to reach the welfare-maximizing level of installed capacity. The first market design regime studied is the canonical capacity market. We build on the previous literature and the design found in Léautier (2016) and Holmberg and Ritz (2020), which relies on the assumption that the capacity market does not have any effect beyond increasing the investment level. Following the approach of our paper, this canonical regime is similar to having a cost allocation regime based on a lump-sum tax. In this case, even when considering the endogenous supply function in the capacity market, the equilibrium of the market design always restores the first-best optimum given the system inefficiencies. Namely, providing that the demand function is intersecting the supply function at the optimal level,[14] the equilibrium price in the capacity market and the suboptimal price in the subsequent markets is always equal to the price in an optimal system without price cap.[15] We

[13]In their seminal paper, Joskow and Tirole (2007) demonstrate that wholesale markets with a price cap cannot lead to the first-best solution. Following this approach, which also serves as a reference model in our paper, Zöttl (2011) developed a theoretical result on investments under *Cournot* oligopoly with discrete investment and a price cap. Using the same model, Léautier (2016) showed that market power from producers could also be a significant cause of underinvestment. He also introduces a capacity market in the benchmark model where producers can exercise market power. This paper serves as our reference for our implementation of the capacity market.

[14]In this paper and unlike Hobbs et al. (2007), we do not analyze the risk of having regulatory errors.

[15]We also demonstrate that this result also holds for other types of inefficiencies. The equilibrium capacity price equals the expected lost revenue with the price cap. In the case of capacity as a public good, the price is equal to the difference between the private value of the investment and its social value.

then investigate the case in which the capacity price impacts the consumers at the margin. In this case, the regime similarly allocates the capacity price as a unitary tax. However, the main difference is that the marginal effect is endogenously determined at the equilibrium because the capacity equilibrium price causes it. We show that the existence of the capacity market indirectly affects the wholesale market by redistributing the different states of the world when the capacity does not bind and bind and by lowering the consumer's surplus. Therefore, we demonstrate that the welfare outcome at the equilibrium under this regime is always lower than under the canonical regime.

We then compare the two capacity cost allocation regimes by including inefficient rationing. When the price cap is reached, the investment availability becomes a public good as the demand becomes inelastic. Due to the impossibility of efficiently rationing consumers, they incur a significant welfare loss.[16] This additional assumption regarding inefficient rationing has significant implications for comparing the two market designs at their respective equilibria. Indeed, under this new assumption, we find that the indirect effect created by allocating the capacity price on a unitary basis is now ambiguous for social welfare. Under canonical model specification, we find that this market design constantly brings more social welfare at equilibrium than the initial allocation regime. This is due to the interaction between each market design's effects on the system and the equilibrium investment level.

As a third step, we extend our analysis to implementing a regime where the regulator allocates the cost based on actual retailers' market shares. It allows us to introduce the ex-post temporality in the current analysis, where the design of capacity markets considers the realized demand and analyzes the effect of retailers' market power in the model. We first show how this design affects at the margin the retailers who play '*à la Cournot*' on the retail market, and then we integrate the new equilibrium into our model with investment decisions and the capacity market. We find that this allocation creates an intermediary outcome between the unitary tax and the lump-sum tax

[16] Using the same initial model Holmberg and Ritz (2020) showed that additional capacity payment is necessary when the system includes the public-good nature of the investments. Indeed, the inadequacy between capacity and consumption generates negative externalities. Hence, to fully internalize the effect of capacity inadequacy, it is necessary to generate an adder on the wholesale price. We also use in this paper the same representation of the public-good nature of the investment. This effect of a price cap is also closely related to the concept of reliability externality described by Wolak (2021).

while having significant redistributional properties. Finally, we also study the effect of the retail market structure on the equilibrium outcomes of the model.[17] Depending on the assumptions with respect to the model parameters and the inefficiencies assumptions, we find that lowering the number of retailers can provide additional social welfare.

Finally, we analyze the case of a capacity market entirely based on the realized demand level. To do so, retailers are obliged to cover the quantity sold on the retail market by buying directly on the capacity market, given a penalty system. We focus on how retailers' individual strategies can form an aggregated demand function in the capacity market, and we analyze the optimal capacity bought by retailers in the capacity market. We find that such an approach for the demand function can provide the optimal level of investment under specific conditions. The market equilibrium under this regime relies on the marginal value a capacity brings to retailers' profit, which also depends on the market structure in the retail market, the consumers' demand function, and the penalty system.

We provide in Section 2 a reminder of the benchmark model that describes investment decisions in generation capacity. We implement the capacity market and build the theoretical supply function in the same section. The different allocation methods are studied in Section 3 and in Section 4. Section 5 provides the analysis of the retailers' participation in the capacity market. To conclude, we discuss possible extensions of the model in Section 6.

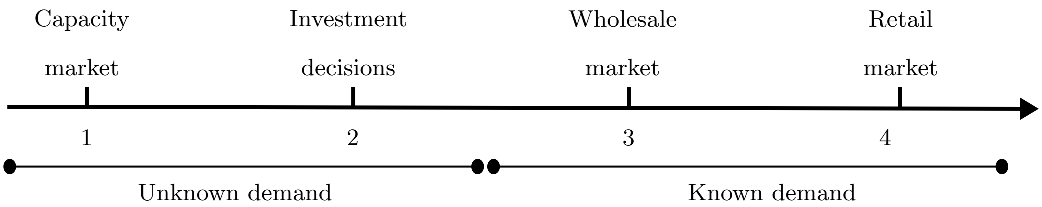
2 Benchmark model with capacity market

2.1 Environment

We consider an initial economic system with three types of agents: producers, retailers, and final consumers. Producers invest in capacities to produce a homogeneous good. They sell the goods to retailers on a wholesale upstream market. Then, retailers resell it to consumers on a downstream retail market.

[17] We do not consider market power on the supply side in our paper, as it is well documented in the literature, see for instance Zöttl (2011) and Leautier (2018) for its effect on investment decision with a price cap, see Léautier (2016) for its effect with a capacity market.

Model stages. We consider a four-stage non-cooperative game. First, producers participate in the capacity market. Second, they choose the level of investment. Third, the wholesale market clears. Fourth, the retail market clears.[18] All decisions during the stages are publicly known and are made simultaneously. We assume the final consumers' demand is uncertain for all agents when making investment decisions. On the other hand, the demand is known when the producers and retailers sell the goods. Those two stages can be interpreted as a repetition of multiple states of the world over a given period (for example, one year), drawn from the distribution (Leautier, 2018). Every agent is to be risk-neutral and maximize expected profit. The game is solved by backward induction.



Producers. We assume perfect competition on the supply side. Producers use a single technology to produce the good. It is characterized by a variable unitary cost c and a fixed unitary investment cost r. We normalized the capacity level, so one unit of capacity allows to produce one unit of the good. The total level of capacity installed after the first stage is k.

Retailers. We allow retailers to be either perfectly competitive or to compete *à la Cournot* to resell the goods to final consumers. Still, they do not behave as an oligopsony in the wholesale market. The imperfect competition is modeled using a finite number of retailers n. We model the retail market as perfectly competitive in Section 3 and 4 to keep the analysis tractable. In Section 5, we introduce the effect of imperfect competition. The use of a finite number is always explicitly indicated. We assume that retailers incur no cost when reselling from the wholesale market to the

[18] As discussed later, the stage order between capacity market and investment decisions does not matter in this model.

retail market apart from the wholesale price. Therefore, perfect competition implies that prices are strictly equal in the wholesale and retail markets.

Demand. The following assumptions characterize final consumers in the retail market. They have the same individual uncertain demand with an aggregate demand function $D(p,t)$, t being the state of the world. t can be understood as the demand level affecting only the intercept of the demand function, not its slope. The demand uncertainty is a random variable characterized by a distribution function $f(t)$ and a cumulative distribution function $F(t)$, which is common knowledge. The inverse demand function is $p(q,t)$, with q the quantity sold on the retail market, such that $D(p(q,t),t) = q$. For convenience, we assume that $p^s(q,t)$ is the price on the wholesale market, and $p(q,t)$ is the price on the retail market. Moreover, the demand function have the following properties[19]: $\forall t \in [0, +\infty)$ (i) $p(q,t) = \dot{p}(q) + \varepsilon(t)$, (*additive demand shock, which implies that* $p_{qt} = p_{tq} = 0$). (ii) $p_t(t) > 0$ (*states of the world are ordered*), $p_{tt}(t) < 0$ (*demand decreases less with higher states*) (iii) $p_q(q) < 0$ (*decreasing inverse demand with respect to* q) $p_{qq}(q) \leq 0$ (*concave inverse demanded that implies decreasing marginal revenue*) (iv) $\lim\limits_{q \to +\infty} p(q,t) < c$ (*prices can be below the marginal cost for some* t). To ensure producers investment in capacities, we need additional conditions: $p(0,t) > c + r \quad \forall t$.

Price Cap. Essential goods are characterized by inefficiencies that prevent the market investment from reaching the first-best economically efficient. Two main reasons private investors do not provide sufficient capacities: (1) the revenue collected on the market is insufficient to cover their production and investment costs, (2) prices do not consider the positive externalities implied by their availability during high demand periods. For the first rationale, we derived the inefficiency that typically characterized essential goods such as electricity: the suboptimality of the wholesale price modeled via a price cap[20]. Further in this paper, we present two other rationales: the public-good nature of capacity during peak-demand states of the world and a concentrated

[19]For most of the functions $f(x,y)$, $f_x(x,y) = \frac{\partial f}{\partial x}(x,y)$, $f_{xx}(x,y) = \frac{\partial^2 f}{\partial x^2}(x,y)$,$f_{xy}(x,y) = \frac{\partial^2 f}{\partial x \partial y}(x,y)$

[20]This modeling approach can represent both an explicit and implicit price cap. In the latter case, political interventions due to the essential nature of the good can artificially alter the price. For instance, when the power system operator needs to carry out technical interventions to avoid system failures. Those policy interventions, such as price caps and non-economic distortions made by a public entity, lead to a *Missing Money* issue that prevents sufficient revenue from being collected to cover costs (Joskow and Tirole, 2007). The effect of price cap regulation was illustrated during the COVID-19 crisis in Italy when the government introduced a 50-cent cap on sales price per mask, which eliminated the incentives to reconvert plants or increase production (Fabra et al., 2021).

retail market represented via retailers' market power. We implement a price cap denoted p^w. To create inefficiencies, the price cap must be binding for some states of the world, so it needs to be below the highest price during the highest demand period; $p^w < \lim_{t \to \infty} p(0,t)$. However, to allow for investment, we also need the price cap above the total unitary cost: $p^w > r + c$.

Closed-form application.

Example 1. Continuous Time Suppose the final consumers' demand function is linear, and the uncertainty comes from the intercept. We define the inverse demand function as follows: $p(q,t) = a(t) - bq$. Where $a(t)$ is the uncertain intercept such that $a(t) = a_0 - a_1 e^{-t}$. We assume that t follows an exponential distribution, which is characteristic of goods with peak demand: $f(t) = \lambda_1 e^{-\lambda_1 t}$ with $\lambda_1 \in (0, 1]$.

Example 2. Discrete Time Suppose that there are only two states of the world. A high demand state, such as the (linear) demand function of final consumers, is equal to $p^h(q) = a_0 - bq$, and a low demand state such as the demand function is $p^b(q) = a_0 - a_1 - bq$. We denote the probability of having a low demand state as θ such as $\theta \in (0, 1)$, with $1 - \theta$ the probability of having a high demand state.

2.2 Market Equilibrium with a Capacity Market

We now describe the equilibrium of the game that consists of a series of equilibria for each stage : (i) the retail market, (ii) the wholesale market, (iii) investment decisions, and (iv) the capacity market. For now, we assume only a direct supply-side effect of the capacity market on the game equilibrium via an increase in the producers' profit. In the rest of the paper, the analysis of the different designs follows the same backward induction but takes into account the link between procurement design.

Fourth stage - Retail market. We assume that symmetric retailers can act strategically '*à la Cournot*' in the retail market, and they take the wholesale price as given. The retailer's profit made on the retail market is $\pi_i^r(t) = q_i(p(q,t) - p^s(q,t))$. The first-order condition gives the equality between the marginal revenue and the marginal cost. In case of imperfect competition, the inverse demand function of retailers on the wholesale market is a downward rotation at the intercept of the final consumer demand function $p^s(q,t) = p(q,t) + \frac{q}{n}p_q(q)$. When the retail market is perfectly competitive, we have straightforwardly: $p^s(q,t) = p(q,t)$. For notation clarity, we assume first perfect competition and use $p(q,t)$ as general notation for demand.

Third stage - Wholesale market. Producers know the final consumer demand at this stage, so the retailers' inverse demand function is certain. The price is determined by the total investment level k chosen during the first stage. We assume perfectly competitive producers, so when k is not binding, the price is equal to the marginal cost c (*off-peak periods*). When k is binding, the price rises above marginal to ensure supply equals demand (*on-peak periods*). We denote $t_0(k)$ the first state of the world when capacity is binding, that is, when the price at the capacity level is equal to the marginal cost: $t_0(k) = \{t : p(k,t) = c\}$. We also define $q_0(t)$ as the quantity bought by final consumers when the retail price is equal to the marginal cost, such that $q_0(t) = \{q : p(q,t) = c\}$. During off-peak periods, when $t_0(k) \geq t$, the price on the wholesale market is the marginal cost c, and the price on the retail market is equal to $p(q_0(t))$. During peak periods, when $t > t_0(k)$, the demand function determines the price with $p(k,t)$.

$$q^s(t) = \begin{cases} q_0(t) & \text{if} \quad t \in [0, t_0(k)] \\ k & \text{if} \quad t \in [t_0(k), +\infty] \end{cases} \qquad\qquad p^s(t) = \begin{cases} c & \text{if} \quad t \in [0, t_0(k)] \\ p(k,t) & \text{if} \quad t \in [t_0(k), +\infty] \end{cases}$$

Second stage - Investment decisions. At this stage, final consumer demand is unknown, and so is the wholesale and retail price. The expected profit of producers is defined as the sum of the expected profit made on the wholesale market and, if implemented, the realized profit on the capacity market minus the investment cost :

$$\Pi^s(k) = \Pi^w(k) + \Pi^c(k) - rk = \int_t q^s(t)(p^s(t) - c)dF(t) + p^c(k)k - rk$$

The market equilibrium in terms of investment decisions with a perfect competitive framework is given by following the first-order condition:

$$\frac{\partial \Pi^w}{\partial k}(k) + p^c(k) - r = 0$$

The crucial term to determine the market equilibrium is the net expected marginal revenue made on the wholesale market. We define $\phi(k)$ for this value for general notation, and it is found by taking the derivative of the wholesale profit: $\phi(k) = \frac{\partial \Pi^w}{\partial k}(k)$. During off-peak periods, producers are perfectly competitive, and prices equal marginal cost; therefore, the marginal revenue is null. When the capacity is binding, it is the difference between the wholesale price and the marginal production cost. The following equation formally defines this rent under the full efficiency assumption (with no price cap):

$$\phi_0(k) = \int_{t_0(k)}^{+\infty} \underbrace{p(k,t) - c}_{\text{on-peak } k \text{ rent}} \, dF(t) \tag{1}$$

We turn now to the framework with a price cap. We introduce a second threshold $t_0^w(k)$. It is the first state of the world when the price cap is binding, that is, when the price at the capacity level is equal to the price cap: $t_0^w(k) = \{t : p(k,t) = p^w\}$. We also define $q_0^w(t)$ as the quantity bought by retailers (or consumers under perfect competition) when the price is equal to the price cap, such that $q_0^w(t) = \{q : p(q,t) = p^w\}$.

$$\phi_0^w(k) = \int_{t_0(k)}^{t_0^w(k)} \underbrace{p(k,t) - c}_{\text{on-peak } k \text{ rent}} \, dF(t) + \int_{t_0^w(k)}^{+\infty} \underbrace{p^w - c}_{\text{on-peak } p^w \text{ rent}} \, dF(t) \tag{2}$$

The conditions on p^w relatively to the marginal cost c ensure that $t_0^w(k) > t_0(k)$.[21]

First stage - Capacity Market.

[21] Under our framework of a single producing technology and without market power on the supply side, the price cap is only binding during on-peak periods. See Zöttl (2011) and Leautier (2018) for a study of price caps with market power and multiple technologies.

We turn to the definition of the equilibrium p^c of the capacity market. For such equilibrium to exist, we make a market condition following Léautier (2016): There are no short sells, meaning that producers cannot sell more capacity than they own. The existence of the capacity market and the no short-sell assumption leads to the following observations: (i) Decision timing does not matter given our current setting: results still hold if the capacity market is set before or after the investment decision as long as it is before the final consumers' demand is known. (ii) It is optimal for producers to offer all their capacities if the first condition holds.[22]

The market equilibrium is found via the intersection between the demand and supply functions offered by producers. For now, we remain agnostic on determining the demand function. Except for 5.2, the demand function is assumed entirely exogenous in the sense that it is determined by a Social Planner that seeks to maximize welfare and consists in a vertical line.

We build the supply function based on the assumption that producers offer their marginal profit loss associated with the capacity market's participation. The common approach in the literature represents the cost of investing beyond the optimal capacity level. However, to our knowledge, this is the first time a supply function in a capacity market is directly modeled using the benchmark framework. As we assume perfect competition in the wholesale market compared to Léautier (2016) and Zöttl (2011), capacity choices have no marginal effect on the rent. Indeed, the rent appears only when total capacity is constraining under perfect competition.[23] The full profit with a capacity market for a producer is $\Pi^s(k) = \phi(k)k - rk + p^c(k)k$. Under perfect competition, the first-order condition gives $\phi(k) - r + p^c(k) = 0$. Therefore, the capacity market's supply function equals the marginal cost associated with the deviation from the market investment level $\bar{k}$, which would have been made without the capacity market.

[22] The intuitions behind the extension of Léautier (2016)'s proposition in the paper are as follows: without a direct link between the quantity exchanged in the capacity market and the investment level (i.e., short sell condition); the former does not alter the producer's marginal profit with respect to the latter. Hence, the capacity market does not have any effect on the investment decision. For observation (i), the result is straightforward as we do not include any specification in terms of investment dynamics (e.g., the time to build the investment) and information structure (e.g., the uncertainty of the demand level can reduce when the investment decision is closer to the wholesale market). For observation (ii), the proof relies on the result that the supply and demand function's outcome in the capacity market leads to a unique symmetric equilibrium as the profit function is also concave.

[23] Under imperfect competition on the supply side, the rent also exists due to market power and can appear before the total capacity is binding.

Definition 1. We denotes the supply function $X(k)$ and the inverse supply function $X^{-1}(p^c)$ such that $X^{-1}(X(k)) = k$. The supply function on the capacity market is defined as follows:

$$X(k) = \begin{cases} 0 \quad \text{if} \quad k \leq \bar{k} \\ r - \phi(k) \quad k > \bar{k} \end{cases} \tag{3}$$

With $\bar{k}$ the market equilibrium given by $r = \phi(\bar{k})$.

Below $\bar{k}$, the marginal cost is positive, and the supply is null. Indeed, as the wholesale market's profit function is concave, any marginal revenues on the left side of the optimum are above the marginal cost of r. The marginal revenue is below the marginal cost on the right side of the optimal investment level. Therefore, any deviation to the right creates a positive opportunity cost.[24] This approach is particularly relevant as it fully characterizes the effect of different market design regimes in the economic system. In other words, if a regime changes the expected revenue made in the wholesale and retailer market, we can consider its feedback effect on the supply function in the capacity market. For a given demand function, the equilibrium of the capacity market is simply found by equalizing the demand function to the supply function. Figure 1 gives an example of how supply functions are built in the capacity market for different assumptions (without or with a price cap and for different values of the price cap). Note the kinks at the bottom for curves; they represent the level of investment that maximizes expected profits; therefore, on the left, the profit is concave, which implies a null supply function. For sufficiently high values of value k, the capacity level does not bind with positive probability. Hence, no rent is generated on the wholesale market. This explains the convergence towards the marginal investment costs.

Therefore, the equilibrium of the game comprises : (i) a wholesale demand function adjusted for the retailers' market power, if implemented, (ii) a wholesale schedule of prices and quantity for each state of the world, (iii) an investment decision based on the expected wholesale market revenue

[24] Our approach to the supply function in the capacity market is similar to the theory of supply function equilibria where bidders offer a function such that each point on this function maximizes their profit/utility (Klemperer and Meyer, 1989). In our paper, the supply function in the capacity market is built such that each producer is indifferent between providing their investment market equilibrium or any investment on the curve in return for the corresponding capacity price.

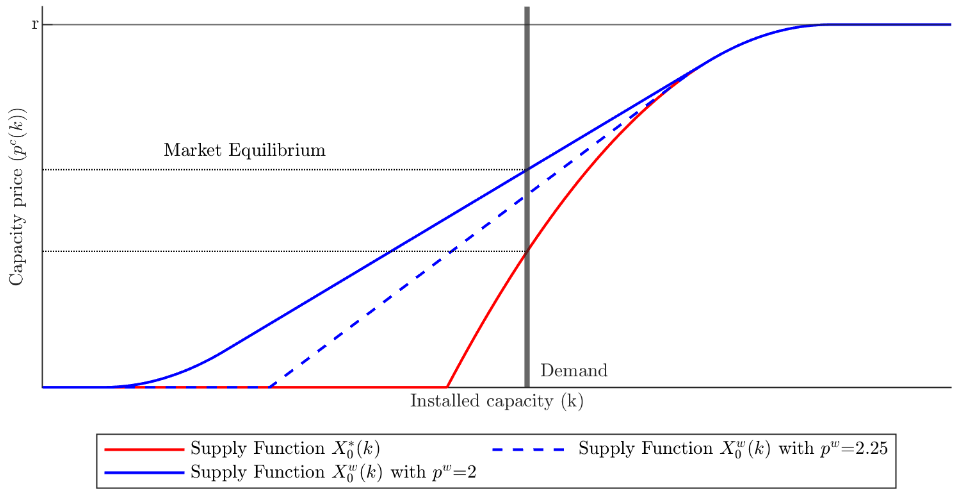


Figure 1: Illustration of the capacity market equilibrium given a demand k with the linear continuous model.

and from collected capacity market revenue (iv) a capacity market equilibrium price coming from a supply function made by producers and a demand function corresponding to a specific level of investment chosen by the social planner.

In the rest of this section, we describe two benchmark equilibria: the first-best level without any form of inefficiency and the second-best market equilibrium with a binding price cap in the absence of a capacity market.

2.3 First-Best without a capacity market

We find the optimal first-best investment level[25] as the value of k that maximizes the expected social welfare without any form of inefficiency. For general notation, we define $W(k)$ as the expected social welfare, comprising the consumer, producer, and retailer's surplus. Under the full efficiency assumption, we defined k_0^* as $k_0^* = \max_k W_0(k)$, with $W_0(k)$ formally defined as follow.

[25] We use the term first-best, socially optimal, and welfare-maximizing interchangeably.

$$W_0(k) = \int_0^{t_0(k)} \underbrace{\int_0^{q_0(t)} (p(q,t) - c)dq}_{\text{off-peak weflare}} \, dF(t) + \int_{t_0(k)}^{+\infty} \underbrace{\int_0^{k} (p(q,t) - c)dq}_{\text{on-peak weflare}} \, dF(t) - rk$$

The maximum k_0^* is found by equalizing the marginal surplus gain from an increase of capacity to the marginal cost :

$$\phi_0(k) = \int_{t_0(k)}^{+\infty} (p(k,t) - c)dF(t) = r \tag{4}$$

Under the initial assumptions, the expected social welfare is concave with respect to the level of investment k, which ensures the existence of a maximum ($\frac{\partial \phi_0}{\partial k} \leq 0$). In the absence of a price cap or any other inefficiency, it is straightforward that the market equilibrium is the first-best solution to maximizing the expected social welfare as the private marginal revenue equals the marginal social revenue.

2.4 Second-best without a capacity market

We now define the market equilibrium with a price cap but without a capacity market. The price cap does not change the social welfare function at the marginal, equal to $W_0(k)$, as it only redistributes surpluses between consumers, producers, and retailers. Without a capacity market, the market equilibrium k_0^w is found by equalizing the expected marginal private profit made on the wholesale market to the marginal investment cost:

$$\phi_0^w(k) = \int_{t_0(k)}^{t_0^w(k)} (p(k,t) - c)dF(t) + \int_{t_0^w(k)}^{+\infty} (p^w - c)dF(t) = r \tag{5}$$

The lemma 1 shows that a price cap in the wholesale market lowers the market investment level and increases inefficiency. We also provide the optimal payment associated with restoring the optimal investment level. It is equal to the expected difference between what should have been the wholesale price and the price cap when it is binding — this is commonly known as the "Missing Money".

Lemma 1. A binding price cap leads to a lower installed capacity than the optimal investment level given by the social welfare maximization: $k_0^w \leq k_0^*$ as well as lower expected welfare: $W(k_0^*) \leq W(k_0^w)$. The optimal capacity payment $z^w(k)$ is :

$$z^w(k) = \int_{t_0^w(k)}^{+\infty} (p(k,t) - p^w) dF(t) \tag{6}$$

Proof. See Appendix □

3 Capacity cost allocation under price cap

In this Section, we study the effects of the allocation of the capacity price on the system equilibrium described previously. From an implementation perspective, assume that a Single Buyer (for instance, the Social Planner) procures all the capacity on the capacity market. Then, the Single Buyer can choose two general regimes to pass through the procurement cost to the consumers [26]: (i) lump sum tax, which boils down to assuming a system equilibrium kept unchanged by the allocation of the capacity price (exogenous design) (i) a variable tax, which increases the price of the good and generates a specific effect on the system equilibrium (endogenous design). This Section proposes a way of solving the new equilibrium and compares the two outcomes with solely missing money inefficiencies created by the price cap. Section 4 extends this analysis by including inefficient rationing.

3.1 Exogenous design

We do not make any assumptions about the identity of the (single) capacity buyer in this paper as it is outside the scope. We assume that this entity forecasts the future expected demand of

[26] We do not differentiate retailers or consumers in this Section as we assume perfect competition.

final consumers, and then it builds the demand function in the capacity market to maximize the expected social welfare. In our analysis, this demand function corresponds to a vertical line equal to the investment level that maximizes the expected social welfare. Indeed, perfect competition on the supply side implies that producers always offer their marginal cost, and the shape of the demand function does not matter. Finally, to balance its budget, it transfers the full purchasing cost to the retailers using an exogenous ratio or directly to consumers via a lump-sum tax. We formally describe this market design regime as follows:

Assumption 1. A single entity chooses a level of investment to buy on the capacity market that maximizes $W(k)$ at a price $p^c(k)$ given by the supply function described in 3. Then, it allocates the full cost $kp^c(k)$ to the retailers or directly to the consumers without any dependence on the expected and realized final demand level.

This assumption corresponds to the traditional approach used in the literature on the capacity market. We call this market design the exogenous regime because (i) the allocation of capacity costs does not depend, for instance, on retailers' realized strategy but rather on exogenous factors such as their past market share (ii) the design does not depend on realized demand for the final good. In other words, this regime only describes the capacity markets' direct effect via the incentive to invest by the capacity price. There is no effect on the final demand because this remuneration is simply a surplus transfer from consumers to producers. This approach's result is that the capacity price equals the optimal payment, allowing the restoration of an optimal level of capacity when the vertical demand function for capacity is calibrated to k_0^*. Whatever the type of inefficiency is considered.[27] This result is described in the following Proposition. It implies that the cost of a capacity market is strictly equal to the transfer necessary to restore the optimal capacity level.

Proposition 1. Under an exogenous allocation market design, the clearing price in the capacity market at the first-best investment level k_0^* given by the supply function $X_0^w(k)$ is always equal to the optimal payment $z^w(k)$ needed to restore efficiency.

[27] We demonstrate in the Appendix that the Proposition can be expanded to the inefficiencies created by retailers' market power and by the public-good nature of capacities.

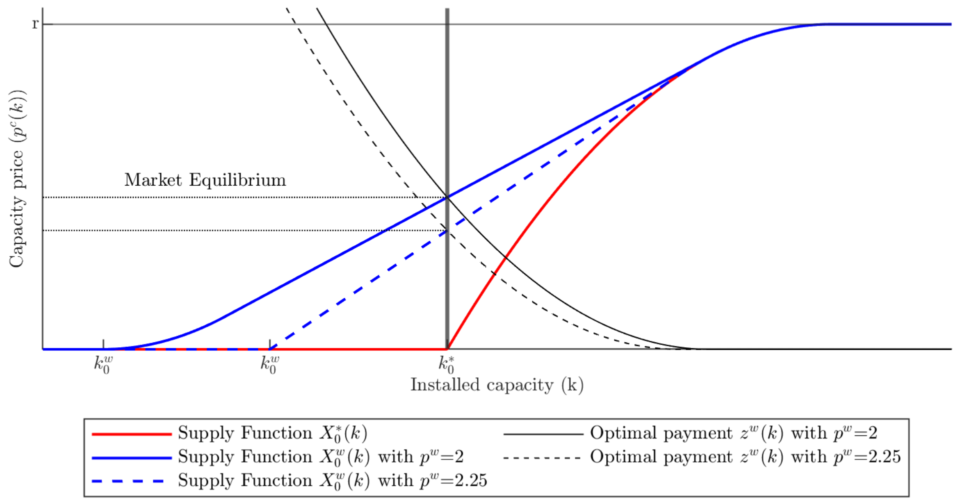


Figure 2: Supply function and equilibrium capacity prices on a capacity market under an exogenous regime with linear demand and exponential distribution.

Proof. See Appendix □

This result highlights the discussion between implementing a price or a quantity instrument to resolve the market inefficiencies or constraints (Weitzman, 1974; Holmberg and Ritz, 2020). We show in this Proposition that the outcome of the capacity market is strictly equivalent to a capacity price set by the regulator defined in equation 6. Under this regime, the exogenous approach is optimal because it gives the right investment level given the inefficiencies.

We illustrate Proposition 1 with the model specification of Example 1. In figure 2, we show two supply functions on a capacity market. When a price cap is introduced, the marginal cost of providing an additional capacity increases, which shifts the supply curve to the left (blue curves), and the market investment level k_0^w decreases compared to the first-best investment level k_0^*. We also include the function associated with the optimal payment as described in equation 6 (black curves). As demonstrated in Proposition 1, the capacity price at the optimal investment level k_0^* equals the optimal payment.

Our model can also provide some comparative statics on the capacity price given the specification of Example 1. For instance, when only the price cap is considered in the model, the results are intuitive and in line with previous works, with the capacity price always being positively impacted by an increase in the demand intercept or by the product costs (variable and fixed). In contrast, the price cap has a negative effect.

3.2 Endogenous allocation

We introduce a new allocation regime for the capacity market. In this case, capacity prices marginally impact the final consumer demand via the allocation of the capacity price. The setting is similar to the previous one, with a single entity forecasting the future expected demand and building the demand function in the capacity market. We formally describe this market design regime as follows:

Assumption 2. A single entity chooses a level of investment to buy on the capacity market that maximizes $W(k)$ at a price $p^c(k)$ given by the supply function described in 3. Then it allocates the full cost $kp^c(k)$ either to the retailers or directly to the consumers such that the new final demand for the good is equal to $D(p,t) = p(q,t) - p^c(k)$.[28]

Compared with the previous setting without this indirect effect, the first previous case can be understood as an increase of the fixed part in a two-part tariff (or a lump sum tax). In contrast, the second case in this subsection can be understood as an increase of the variable part (or a unitary tax). However, the main difference with a price instrument such as a tax is that the capacity price and investment decisions emerge from profit maximization from the producers and the single buyer

[28] This allocation rule is compatible with the budget constraint of the single entity. Another way of expressing this allocation rule would be using the value $p^c(k)\frac{q}{k}$. While this allocation changes the numerical result, it does not impact the fundamental results of this section.

choice of the demand function. Therefore, we formally demonstrate the existence of the indirect effect by repeating the steps of the previous model and using backward induction.

Third stage - Retail market. Let $p^c(k)$ (or p^c for notation clarity) be the capacity price adder for final consumers, identical to a unitary consumption tax as the value is sunk at this period. The final consumer demand function shifts downward with its new value equal to $p(q,t) - p^c(k)$. k is still the quantity bought on the capacity market by the entity at a price $p^c(k)$. We denote $t_1(k)$ and $q_1(k, p^c)$ the new thresholds for respectively the states of the world between on-peak/off-peak periods such that $t_1(k) = \{t : p(k,t) - p^c(k) = c\}$, and the corresponding quantity such that $q_1(t,k) = \{q : p(q,t) - p^c(k) = c\}$. We also define the thresholds for the price cap with $t_1^w(k, p^c)$ the first state of the world when the price cap is biding under the endogenous design, that is $t_1^w(k,p^c) = \{t : p(k,t) - p^c(k) = p^w\}$. We also define $q_1^w(t)$ as the quantity when the price is equal to the price cap, such that $q_1^w(t) = \{q : p(q,t) - p^c(k) = p^w\}$ For now, we assume that the capacity price exists. We formally demonstrate it in the proof of Lemma 3.

Second stage - Wholesale market While the demand is always lower or equal to the initial demand function, the impact on the expected social welfare is not trivial. The Lemma 2 summarizes the main insight and states that the new welfare function is always lower or equal to the exogenous case.

Lemma 2. Allocating the capacity price at the margin only affects the share between on-peak and off-peak periods and the expected surplus size during off-peak periods. Namely, only the occurrence of the two periods $t_0(k)$ and the intersection between the demand function and the marginal cost $q_0(t)$ change, the welfare function becomes:

$$W_1(k, p^c) = \int_0^{t_1(k,p^c)} \int_0^{q_1(t,p^c)} (p(q,t) - c) dq dF(t) + \int_{t_1(k,p^c)}^{+\infty} \int_0^k (p(q,t) - c) dq dF(t) - rk$$

Proof. See Appendix ☐

We can rewrite the equation by showing the initial welfare function without endogeneity: $\Delta W_1(k,p^c) = W_0(k) - W_1(k,p^c)$. With :

$$\Delta W_1(k,p^c) = \int_0^{t_0(k)} \underbrace{\int_{q_1(t,p^c)}^{q_0(t)} (p(q,t)-c)dq}_{\Delta \text{ in surplus}}\, dF(t) + \int_{t_0(k)}^{t_1(k,p^c)} \underbrace{\int_{q_1(t,p^c)}^{k} (p(q,t)-c)dq}_{\Delta \text{ in occurrence}}\, dF(t) > 0$$

The first part of $\Delta W_1(k,p^c)$ represents the loss when it is off-peak periods for both cases (indeed, we have $t_0(k) \leq t_1(k,p^c)$ as lower demand always means a higher chance of being off-peak): the consumers fully support the loss as producers receive the marginal cost. The second part represents the loss when the capacity level is such that it is an off-peak period with the endogenous case and an on-peak for the other case. Therefore, the loss is shared between consumers and producers, the former sustaining a higher price and receiving a lower margin. There is no loss when both cases are in peak periods, as the quantity on the market is strictly equal to the capacity installed. Hence, recovering the capacity cost allocation only during peak periods does not generate a deadweight loss.

We continue the endogenous regime analysis by defining the economic system's main equilibrium variables. While Proposition 2 and Lemma 2 underline the effects of this regime on the regulatory objective function, we show in the following analysis its effect on market equilibrium, namely the outcome in terms of bidding behavior in the capacity market and in terms of prices.

First stage (i) - Investment decisions Producers make their investment decisions based on the expected net revenue, composed of the expected rent and the capacity revenue. The net revenue is similar to the exogenous case, except for the new state of the world thresholds and the wholesale price. It is defined in the following equation.

$$\phi_1^w(k,p^c) + p^c = \int_{t_1(k,p^c)}^{t_1^w(k,p^c)} (\underbrace{p(k,t)-p^c}_{\text{consumer net demand}} - c)dF(t) + \int_{t_1^w(k,p^c)}^{+\infty} (p^w - c)dF(t) + p^c$$

A sufficient condition for market equilibrium in terms of investment decisions is $\frac{\partial p^c}{\partial k} \geq 0$. In that case, the expected profit is concave, and the second order is satisfied. The derivative of the capacity price depends on underlying assumptions we discuss in the next stage. We also formally prove in the Appendix that $\frac{\partial p^c}{\partial k} \geq 0$. Note that at the second-best market equilibrium with a price cap k_0^*, the supply function is identical under the two market design regimes ($\phi_0^w(k_0^w, p^c) = \phi_1^w(k_0^w, p^c)$). Indeed, there is no opportunity cost of being at k_0^w. The capacity price is null, which implies no indirect effect.

First stage (ii)- capacity market When a producer participates in the capacity market, it bids its marginal opportunity cost without the capacity revenue (but it takes into account its indirect effect on the demand) equal to $r - \phi_1^w(k, p^c)$. Therefore, following the previous stage, the equilibrium is defined with the equality $X(k) = r - \phi_1^w(k, p^c)$. Lemma 3 states how the equilibria are found. It underlines the endogenous nature of this regime where the choice of capacity is both on the supply and demand side of a capacity market and has indirect effects. In other words, the endogeneity of the regime also changes the bidding behavior in the capacity market compared to the exogenous case.

Lemma 3. For any values of $k \in [k_0^w, +\infty)$, there exists a value p^c such that we have $X_1(k, p^c) = p^c$. $X_1(k, p^c)$ is the endogenous supply function in the capacity market:

$$X_1(k, p^c(k)) = r - \left(\int_{t_1(k,p^c)}^{t_1^w(k,p^c(k))} (p(k,t) - p^c(k) - c)dF(t) + \int_{t_1^w(k,p^c(k))}^{+\infty} (p^w - c)dF(t) \right) \quad (7)$$

(ii) Moreover, the supply function is always higher under the endogenous regime than under the exogenous regime: $X_1(k) \geq X_0(k)$

Proof. See Appendix □

Figure 3 describes the change in the supply function when considering the indirect effect of the equilibrium capacity price on the consumers.

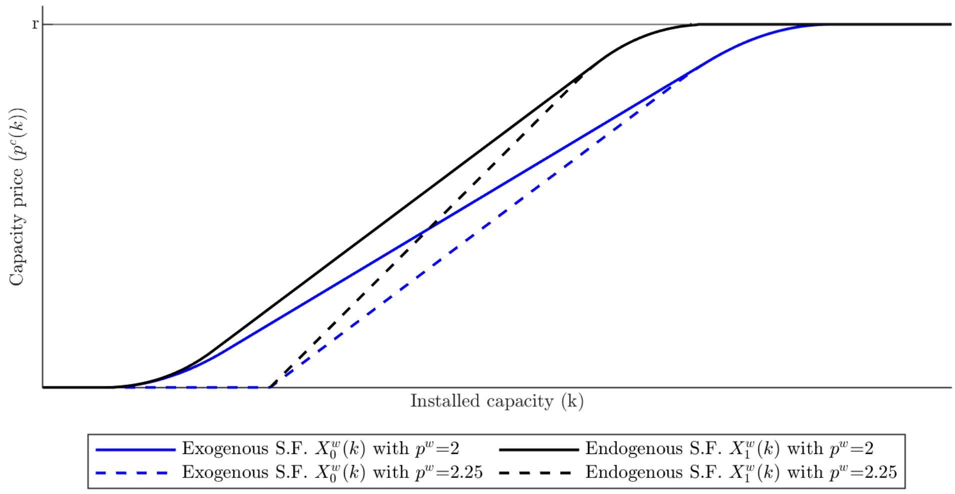


Figure 3: Supply functions in the capacity market for the exogenous and endogenous market regime and for different values of the price cap.

The solution to the fixed-point problem comes from the interaction between the capacity price and the demand in the wholesale market. The proof relies on the observation that for a relatively high level of capacity price, the demand is decreased such that at one point, the capacity and the price cap never bind in expectation. In that case, the existence of a fixed point is always ensured. For other values of the price cap, a fixed point might exist, but the shape of the supply function prevents sufficient conditions from emerging.

We have demonstrated the effect of the endogenous market design on the retail and wholesale market and the condition for an investment market equilibrium and a supply function to be well defined. To fully describe the system equilibrium, we analyze the level of investment bought on the capacity market by the Single Buyer (and therefore installed by producers). The proposition 2 describes the new optimal investment level that maximizes the expected social welfare given this endogenous regime. It has a strong implication as we state that this regime also modifies the objective for the regulator in terms of the final investment level.[29] Moreover, we find that the endogenous regime is always worse than the exogenous regime regarding social welfare.

[29]Therefore, the single buyer also needs to take into account the indirect effect while choosing the demand function.

Proposition 2. (i) The new first-best solution in terms of investment level under the endogenous regime exists if the initial assumptions hold.

(ii) If it exists it solves $k_1^* = \{k : \phi_1(k) = r\}$, with $\phi_1(k)$ defined as follow

$$\phi_1(k) = \int_0^{t_1(k)} \underbrace{\frac{\partial q_1(t,k)}{\partial k} p^c(k)}_{\text{price effect -}} \, dF(t) + \int_{t_1(k)}^{+\infty} \underbrace{(p(k,t) - c)}_{\text{quantity effect +}} \, dF(t)$$

(iii) k_1^* is always lower than the first-best solution under the exogenous level ($k_1^* \leq k_0^*$). The social welfare at the optimal investment level is also always lower than the social welfare at the optimal investment level under the exogenous regime ($W_1(k_1^*) \leq W_0(k_0^*)$).

Proof. See Appendix □

The condition in part (i) of the proposition relates to the concavity of the expected social welfare and the transmission channels of the capacity price in the expected welfare function. The first derivative, represented in $\phi_1(k)$, shows that when the capacity level increases: (a) it decreases the unconstrained quantity $q_1(t,k)$, where at this value the social surplus is equal to the capacity price ($p(q_1(t,k),t) - c = p^c(k)$), this term is therefore negative, (b) it generates an additional surplus during on-peak periods which is equal to $p(k,t) - c$, which is positive, (c) to invest in an additional surplus, one need to sustain the investment cost r. All other marginal effects cancel each other. This marginal value needs to decrease in k to induce a concave expected social welfare. In the proof, we compute the second derivative; we show that the effect at the limits of the respective integrals of $\phi_1(k)$ are easily found to be negative, as well as the marginal change of (b), where a decreasing demand function is sufficient. However, the marginal effect on the equilibrium quantity $q_1(t)$ is more complex. It is directly relative to the shape of the demand function concerning the uncertainty and the distribution function $f(t)$. This is ensured by having $\frac{\partial}{\partial k}\left[\frac{\partial q_1(t,k)}{\partial k} p^c(k)\right] \leq 0$, that is the marginal loss sustained during off-peak periods, accounting for the indirect effect of the endogenous regime is decreasing. A convex capacity market supply function is a sufficient condition to have $\frac{\partial^2 q_1(t,k)}{\partial k^2}$ to be decreasing. The convexity of the capacity market supply function is ensured by having the expectation factor $\Delta F_1(k) = \int_{t_1(k)}^{t_1^w(k)} dF(t)$ to be increasing with k that is: $\frac{f(t_1^w(k))}{p_t(t_1^w(k))} \geq \frac{f(t_1(k))}{p_t(t_1(k))}$. It has different implications depending on the assumptions regarding $f(t)$ and the effect of t on the inverse demand function. Following the classic assumption of Example 1, the exponential distribution implies that $f(t_1^w(k)) < f(t_1(k))$, it is less likely for the price cap to binds compared to the capacity k. On the other hand, if $p_{tt} < 0$, that is, the inverse demand function increases less for higher states of the world (at the margin), it implies that $t_1^w(k)$ increases more than $t_1(k)$, as it is more likely to binds (at the margin). If the second effect dominates the first, then $\frac{f(t_1^w(k))}{p_t(t_1^w(k))} \geq \frac{f(t_1(k))}{p_t(t_1(k))}$. The policy result of (iii) stems from the analysis of the derivative of $\Delta W_1(k)$ with respect to the level of investment k, which is always positive. The condition for the existence of a first-best investment level is sufficient and implies that $\phi_1(k)$ is decreasing with respect to k (i.e., the expected social welfare $W_1(k)$ is concave).

We conclude this section by studying the capacity market equilibrium price. From Lemma 3 and Proposition 2, we know that implementing the endogenous regime (i) lowers the investment level and (ii) increases the supply function on the capacity market. It implies the regime has an ambiguous effect on the capacity market equilibrium: endogenous capacity prices can be higher or lower than exogenous capacity prices, even though the capacity quantity is always lower. It has significant policy implications as the capacity price is politically sensitive as it can make up a significant part of producers' revenue and consumers' bills. Figure 4 shows how the indirect effect can lead to opposite results in terms of equilibrium prices: the capacity price raises the supply function but also decreases the investment level that maximizes the expected social welfare. Therefore, the new equilibrium price can either be above or below the exogenous equilibrium.

We define in Lemma 4 Sufficient condition such that the capacity price is always higher under the endogenous regime compared to the exogenous one. In claim 1, we provide numerical-based results under the exponential and discrete model specifications.

Lemma 4. The capacity price at the first-best investment level is higher under the endogenous regime compared to the exogenous regime if the following condition holds

$$\frac{1}{2}\sum_i \left(\Delta F_i(k_i^*)\right)\left(p^c(k_1^*) - \Delta p(k^*)\right)$$

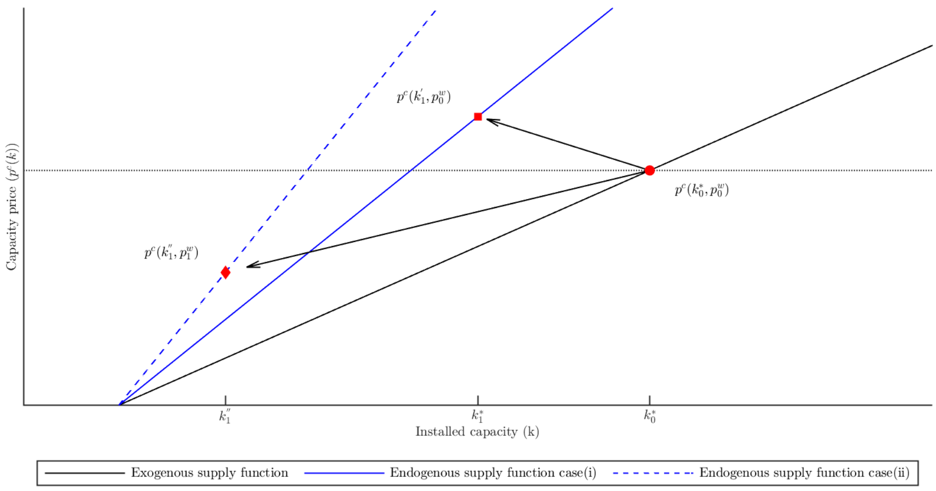


Figure 4: Caption

With $\Delta F_i(k_i^*) = \int_{t_i(k_i^*)}^{t_i^w(k_i^*)} dF(t)$ and $\Delta p(k^*) = p(k_1^*, t) - p(k_0^*, t)$

Proof. See Appendix $\square$

The expression is positive if the endogenous capacity price $p^c(k_1^*)$ is higher than the difference between the deterministic part of the inverse demand function. Note that the first part of the model is always positive and represents the average conditional expectation for any state of the world when the capacity binds but not the price cap. The intuition of the proof is based on the relationship between the expected wholesale price producers receive and the sum of the variable and fixed costs. At the first-best investment level for both market design regimes, the sum of the capacity price and the expected wholesale price is equal to the marginal costs $(c+r)$. As the latter are constants, it is sufficient to study the value of the expected wholesale price to conclude the value of the capacity price at the first-best investment level for the two regimes. The previous analysis shows that the indirect effect induced by the endogenous market design modifies the demand and, by extension, the expected prices in the wholesale market. The general condition in Lemma 4 is sufficient. The expected prices in the wholesale market are always lower under the endogenous market design, hence ensuring a higher capacity price.

Close-form answers need additional conditions, notably on (i) the demand function (ii) the uncertainty distribution. However, except for the discrete example, the endogenous nature of the capacity price prevents having a clear close-form expression of the conditions in Lemma 4, even for the linear demand. Numerical simulation shows that the chosen example always leads to a higher capacity price under the endogenous regime.

Claim 1. Under the exponential and discrete distribution, the condition in Lemma 4 always holds. That is, the equilibrium capacity price under the endogenous regime is always higher than under the exogenous regime.

4 Capacity cost allocation with inefficient rationing

4.1 Inefficient rationing and Second-Best

We now introduce the public-good nature of capacity during peak demand via inefficient rationing. In this section, we use this rationale to revise the comparison between the endogenous and exogenous designs for capacity markets. In section 5, we will also keep the inefficiency to analyze other options in the design of capacity markets.

When binding at the price cap level, the price-elastic demand becomes inelastic.[30] Therefore, we face the same rationing problem as in the literature with limited production capacities and inelastic consumers (see, for instance, Joskow and Tirole (2007)).[31] The absence of efficient discrimination between consumers with a heterogeneous willingness to pay implies that investment availability is a public good when the price cap is binding. Therefore, it is underprovided by producers when they make their investment decisions. The literature describes the cost of involuntary rationing

[30] The introduction of retailers into the model in the rest of the paper does not change the intuition. At a price p^w, the *Cournot* competition between the retailers incites them to ask for an equilibrium quantity above the investment value. Then we assume that a regulated entity ration the retailers, such as their final profit and realized sales, are physically constrained by the investment level.

[31] This inefficiency is associated with the existence of a *Missing Market* issue under which producers consider their revenue insufficient to invest optimally (Newbery, 2016). This can be caused by hedging markets being incomplete (De Maere d'Aertrycke et al., 2017), or because of externalities associated with the public-good nature of investment and consumption choices (Holmberg and Ritz, 2020), innovation spillovers, and climate change.

in various ways. Joskow and Tirole (2007) shows that it depends on whether the rationing is anticipated or not. Leautier (2018) finds that the effect of involuntary rationing can be different if it impacts the expected demand level. From a modeling perspective, Holmberg and Ritz (2020) uses a general function $J(.)$ to represent this negative externality. The function depends on the delta between the quantity bought at a price equal to the price cap and the investment level. For general notation, we note this cost $M(k)$, defines as follow :

$$M(k) = \int_{t^w(k)}^{+\infty} J(t,k)dF(t) \tag{8}$$

With Δk a function of the difference between the installed capacity k and the quantity bought by retailers at the price cap q^w ($q_0^w(t)$ or $q_1^w(t)$ depending on the chosen regime). To illustrate the inefficient rationing cost, we use the following assumption.

Assumption 3. Suppose that consumers sustain an additional cost proportional to the share of consumers selected indifferently who are forced to stop consuming based on their expected surplus. The expected cost is equal to :

$$M(k) = \int_{t^w(k)}^{+\infty} \frac{q^w - k}{q^w} \int_0^{q^w} (p(q,t) - p^w)dqdF(t) \tag{9}$$

In this case, the function $J(t,k)$ can be decomposed into two components: the rationing ratio $\frac{q^w(k)-k}{q^w}$ and the consumer welfare at the quantity asked at the price cap (and excluding capacity payments) $\int_0^{q^w}(p(q,t) - p^w)dq$. This example resembles the rationing model used in (Léautier, 2014). can also be interpreted as follows. Assume that there exists a continuum of consumers such that each point on the inverse demand function $p(q,t)$ represents its marginal willingness to pay for the good. In that case, inefficient rationing implies that each consumer sustains the same cost proportionally to its marginal willingness to pay. This illustration leads to $J(k, t^w) = 0$. Without any difference between the capacity and quantity values, inefficient rationing implies no cost. Regarding the sign of the cost and its derivatives, it seems natural to have $\frac{\partial M(.)}{\partial k} \leq 0$, such that the closer the capacity level is to q^w, the lower is the cost. The second derivative of the expression also matters. For instance, Holmberg and Ritz (2020) assumes a convex function : $\frac{\partial M(.)}{\partial k} \geq 0$. However, as discussed below, our assumption is insufficient to satisfy this condition.

We start with extending the benchmark model with inefficient rationing, which also describes the exogenous regime as it does not generate indirect effects. Then, we compare the outcomes under both regimes. First, we express the expected welfare under the exogenous design with inefficient rationing:

$$
\begin{aligned}
W_0^{bo}(k) &= W_0(k) - \int_{t^w(k)}^{+\infty} \frac{q^w - k}{q^w} \int_0^{q^w} (p(q,t) - p^w) dq dF(t) \\
&= \int_0^{t_0(k)} \overbrace{\int_0^{q_0} (p(q,t) - c)\, dq}^{\text{off-peak welfare}} dF(t) + \int_{t_0(k)}^{t_0^w(k)} \overbrace{\int_0^{k} (p(q,t) - c)\, dq}^{\text{on-peak } k \text{ welfare}} dF(t) \\
&+ \int_{t_0^w(k)}^{+\infty} \underbrace{\int_0^{k} (p^w - c)\, dq}_{\text{producer surplus}} dF(t) + \underbrace{\int_{t_0^w(k)}^{+\infty} \frac{k}{q^w} \int_0^{q_0^w(t)} (p(q,t) - p^w)\, dq}_{\text{consumer surplus - } M_0(k)} dF(t) - rk
\end{aligned}
$$

The terms of the second line represent the expected social welfare for any state of the world where the price cap is not binding. The first term in the third line is the producers' expected revenue, as we assume no rationing cost on the supply side. Finally, the last term on the third line is the expected welfare net of the rationing cost and the producer revenue. Note that this differs from the consumer surplus. It does not consider the transfer between consumers and producers from the capacity market, as it is neutral from a social welfare consideration. Compared to the initial inefficiency of a price cap, the social cost of rationing directly affects the social welfare function. On the other hand, producers' expected marginal rent collected on the wholesale market remains unchanged when we include inefficient rationing, which only affects consumers' welfare.[32]

The absence of direct supply-side effects of inefficient rationing allows us to study our environment's last stage directly. That is, how the single buyer chooses the level of investment in the capacity market. We denote k_0^{bo} the optimal level of investment that maximizes the expected social welfare, such that $k_0^{bo} = \{k : \phi_0^{bo}(k) = r\}$, with $\phi_0^{bo}(k)$ as usually defined as the marginal expected social welfare with respect to the investment level k :

[32] Some authors do include those costs in the producer profit, using a fixed reputational cost (Llobet and Padilla, 2018) or a market shutdown during which producers also lose profit (Fabra, 2018)

$$\phi_0^{bo}(k) = \int_{t_0(k)}^{t_0^w(k)} (p(k,t) - c)\ dF(t) + \int_{t_0^w(k)}^{+\infty} (p^w - c)\ dF(t) + \int_{t_0^w(k)}^{+\infty} \underbrace{\frac{1}{q_0^w(t)} \int_0^{q_0^w(t)} p(q,t) - p^w\ dq}_{\Delta \text{ in net consumer surplus}} dF(t)$$

The first and second parts of $\phi_0^{bo}(k)$ are common to $\phi_0(k)$. An increase of k (1) allows a marginal gain when the capacity is binding for consumers and producers, and (2) allows an additional rent for the producers when the price cap is binding. The last term represents consumer gains when the price cap binds the net of the marginal rationing cost. The assumption with respect to the expression of the cost allows having concave expected welfare as $\int_{t_0^w(k)}^{+\infty} \frac{1}{q_0^w(t)} \int_0^{q_0^w(t)} p(q,t) - p^w\ dqdF(t)$ is decreasing with k via $\frac{\partial t_0^w(k)}{\partial k} > 0$.

The following Lemma concludes on the difference between the initial inefficiency caused by a price cap and the consequences of inefficient rationing. The optimal payment to restore the first-best solution equals the marginal value of an additional capacity for the system, which decreases the cost of involuntary rationing.

Lemma 5. When the price cap induces involuntary rationing, the inefficiency is greater than with voluntary rationing. The optimal investment level is greater $k_0^{bo} \geq k_0^*$, and the expected social welfare at the optimum is lower $W(k_0^*) \geq W_0^{bo}(k_0^{bo})$. The optimal capacity payment $z_0^{bo}(k)$ is : $z^{bo}(k) = -\dfrac{\partial M(k)}{\partial k}$

Proof. See Appendix □

We now turn to analyzing the model under the endogenous regime given the rationing cost. We first discuss the differences with the exogenous case, and then we describe the implications for the regulator. The expected welfare function under the endogenous regime is equal to the following:

$$W_1^{bo}(k) = \int_0^{t_1(k)} \int_0^{q_1(t,k)} (p(q,t) - c)\, dqdF(t) + \int_{t_1(k)}^{t_1^w(k)} \int_0^k (p(q,t) - c)\, dqdF(t)$$
$$+ \int_{t_1^w(k)}^{+\infty} \int_0^k (p^w - c)\, dqdF(t) + \int_{t_1^w(k)}^{+\infty} \underbrace{\frac{k}{q_1^w(t,k)} \int_0^{q_1^w(t,k)} (p(q,t) - p^w)\, dq}_{\text{consumer surplus - } M_1(k)} dF(t)$$

We denote k_1^{bo} the optimal level of investment such that $k_1^{bo} = \{k : \phi_1^{bo}(k) = r\}$, with ϕ_1^{bo} again defined as the marginal expected social welfare with respect to the investment level k :

$$\phi_1^{bo}(k) = \int_0^{t_1(k)} \frac{\partial q_1(t,k)}{\partial k} p^c(k) dF(t) + \int_{t_1(k)}^{t_1^w(k)} (p(k,t) - c)\, dF(t) + \int_{t_1^w(k)}^{+\infty} (p^w - c)\, dF(t) +$$
$$\int_{t_1^w(k)}^{+\infty} \frac{1}{q_1^w(t,k)} \int_0^{q_1^w(t,k)} (p(q,t) - p^w)\, dq - \frac{k}{(q_1^w(t,k))^2} \frac{\partial q_1^w(t,k)}{\partial k} \int_0^{q_1^w(t,k)} p(q,t) - p(q_1^w, t) dqdF(t)$$

The first line is common to $\phi_1(k)$. An increase of k (1) increases the capacity price, which decreases the expected surplus during off-peak periods, (2) allows a marginal gain for consumers and producers when the capacity is binding, and (3) allows an additional rent for producers when the price cap is binding. The second line represents the interaction between the consumer surplus (without the capacity price) and the rationing cost. We develop below the derivative of $M_1(k)$ with respect to k

$$\frac{\partial M_1(k)}{\partial k} = \int_{t_1^w(k)}^{+\infty} \underbrace{\frac{-q_1^w(t,k) - k \frac{\partial q_1^w(t,k)}{\partial k}}{(q_1^w(t,k))^2}}_{\Delta \text{ in rationing ratio}} \int_0^{q_1^w(t,k)} (p(q,t) - p^w)\, dq - \underbrace{\frac{q_1^w(t,k) - k}{q_1^w(t,k)}}_{\text{price effect}} p^c(k) dF(t)$$

The marginal consumer surplus when the price cap is binding has a similar interpretation as the marginal consumer when the capacity does not bind (first term in $\phi_1^{bo}(k)$). It is equal to $\int_{t_1^w(k)}^{+\infty} \frac{\partial q^w(t,k)}{\partial k} p^c(k) dF(t)$ and corresponds to the negative price effect of the endogenous regime. It comes from the fact that we base our rationing cost on the quantity demanded by consumers,

$q_1^w(t,k)$, before sustaining the cost of being rationed. This integral minus the expression above equals the second line of $\phi_1^{bo}(k)$.

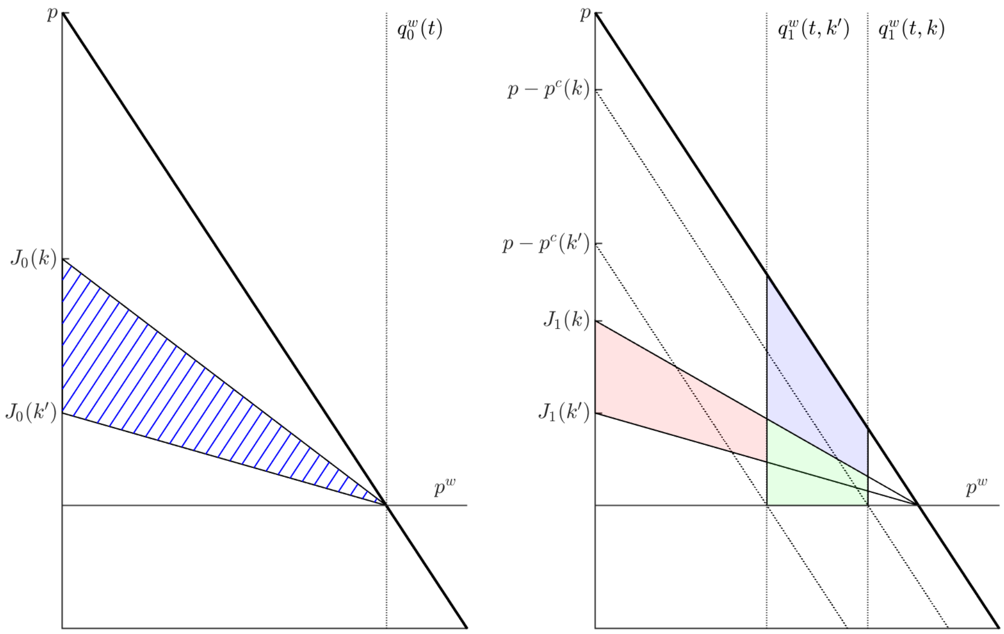


Figure 5: Change in surplus without (left panel) and with inefficient rationing (right panel) under the endogenous regime.

We illustrate the variation of the rationing costs due to the variation of the investment level k with $k > k'$ in Figure 5. The first panel shows the delta between the rationing costs for the exogenous case. The initial rationing cost is the integral (up to q_0^w) below $J_0(k)$ while the new cost is the integral below $J_0(k')$. Therefore, the delta is the hatched area. Note the proportionality of the rationing cost via the downward rotation of the inverse demand function $p(q,t)$. The second panel represents the endogenous case. First, note that the quantity demanded by consumers at the price cap (i) is based on the inverse demand adjusted by the price cap $p(q,t) - p^c(k)$ and not on the inverse demand $p(q,t)$, (ii) decreases with the increase of the investment level, due to the increase the capacity price. The red zone represents the first term in $\frac{\partial M_1(k)}{\partial k}$ and corresponds to the change in the proportional ratio $\frac{q_1^w(t,k)-k}{q_1^w(t,k)}$. This is a similar effect represented in the hatched area of the first panel. The green zone represents the change in the rationing cost due to the indirect effect of the capacity price on the demand function. It is the second part of the derivative $\frac{\partial M_1(k)}{\partial k}$. By increasing the price, the capacity market decreases the demand and the rationing cost size. Finally, the sum of the blue and green zones represents the decrease in consumer surplus (without capacity cost) due to the decreases in the quantity demanded at the price cap (in the Appendix, we show that $\frac{\partial q_1^w(t,k)}{\partial k} \leq 0$). The assumption concerning the form of the rationing costs leads to a net total effect that is always positive (red zone - blue zone), as illustrated in the second line of $\phi_1^{bo}(k)$. The following Lemma describes a sufficient condition for the existence of an endogenous regime equilibrium. That is, the expected social welfare is concave [33].

Lemma 6. If the following condition holds $-\frac{\partial}{\partial k}\left[\int_{t_1^w(k)}^{+\infty} kF(t)\right] \geq \int_{t_1^w(k)}^{+\infty} F(t)$ Then a maximum of the expected social welfare $W_1^{bo}(k)$ exists.

The condition supposes that the conditional expectation decreases sufficiently fast. When the level of capacity increases, it decreases the occurrence of states of the world when the price cap binds. It states this marginal decrease of occurrence at the capacity level $(-kf(t_1^w(k))\frac{\partial t_1^w(k)}{\partial l})$ should be higher than the marginal expected gain of capacity during those states of the world $(\int_{t_1^w(k)}^{+\infty} 1dF(t))$ by a factor of two. From an economic perspective, it ensures that when expanding the capacity level, the gains from reducing the inefficient rationing cost do not increase with the capacity level.

4.2 Exogenous vs. Endogenous regime with inefficient rationing

The main difference between the two regimes is that the capacity price indirectly affects q^w, which now depends on the level of investment. This has significant implications for the comparative statistics between the two regimes. We start by expressing the difference in terms of welfare $\Delta W_1^{bo}(k) = W_1^{bo}(k) - W_0^{bo}(k)$:

[33]Holmberg and Ritz (2020), while not providing a specific form of $M(k)$, assume a convex rationing cost, which is a sufficient condition for the concavity of expected welfare. In our specification, this is not the case.

$$
\begin{aligned}
\Delta W_1^{bo}(k) = & - \int_0^{t_0(k)} \int_{q_0(t)}^{q_1(t,k)} (p(q,t)-c)dqdF(t) - \int_{t_0(k)}^{t_1(k)} \int_{q_0(t)}^{k} (p(q,t)-c)dqdF(t) \\
& \underbrace{- \int_{t_0^w(k)}^{t_1^w(k)} \int_{k}^{q_0^w(t)} (p(q,t)-p^w)dqdF(t) - \int_{t_1^w(k)}^{+\infty} \int_{q_1^w(t,k)}^{q_0^w(t)} (p(q,t)-p^w)dqdF(t)}_{\Delta \text{ in welfare due to price effect}} \\
& + \underbrace{\int_{t_0^w(k)}^{t_1^w(k)} J_0(t,k)dF(t)}_{\text{Avoided rationing cost}} + \underbrace{\int_{t_1^w(k)}^{+\infty} J_0(t,k) - J_1(t,k)dF(t)}_{\Delta \text{ in rationing cost}}
\end{aligned}
$$

The first and second lines represent the endogenous regime's negative effect presented in section 3. However, the rationing cost specification is based on the quantity consumed at the price cap. Hence, with rationing, the endogenous regime also indirectly affects periods during which the price cap binds (second line) and not only on periods off-peak periods (first line). In the third line, the first term stands for the lower occurrence of periods during which the price cap is binding due to the lower demand. In that case, welfare trades a rationing cost against a lower welfare corresponding to the first term in the second line. The second term represents the change in rationing cost due to a change in the quantity consumed at the price cap. We express this term $\Delta J(t,k) = J_0(t,k) - J_1(t,k)$ below after rearrangement :

$$
\Delta J(t,k) = k \left(\frac{1}{q_1^w(t,k)} - \frac{1}{q_0^w(t)} \right) \int_0^{q_1^w(t,k)} (p(q,t)-p^w)dq + \frac{k - q_0^w(t)}{q_0^w(t)} \int_{q_1^w(t,k)}^{q_0^w(t)} (p(q,t)-p^w)dq
$$

We have $\Delta J(t,k) > 0$ from the observations that $q_1^w(t,k) < q_0^w(t)$ due to the negative price effect on the quantity of the endogenous regime. Therefore, for similar states, the rationing cost is always lower under the endogenous regime than in the exogenous case. However, it does not render the comparative statics of the equilibrium straightforward, especially when describing (i) the ranking between the equilibrium investment level and (ii) the ranking between the equilibrium welfare.

We start with the ranking between the different investment levels. Section 3 showed that we always have a lower investment level under endogenous regime $k_1^* \geq k_0^*$ compared to the (first-best) exogenous regime under a sole price cap inefficiency. With inefficient rationing, Lemma 5 showed that the investment level is higher than the first best: $k_0^* \leq k_0^{bo}$. However, due to the opposite effects an endogenous regime exhibits between the negative price effect and the decrease of rationing costs, the ranking between k_1^{bo} and the first-best k_0^* is a priori unclear. Proposition 3 provides a ranking between the investment level under the linear assumptions.

Proposition 3. Under the model specifications, there is a unique ranking between the investment equilibrium such that: $k_1^* \leq k_0^* \leq k_1^{bo} \leq k_0^{bo}$

Proof. See Appendix □

Previous analysis shows that introducing rationing costs increases the investment level, and having an endogenous regime (excluding rationing) decreases it. Hence, the ranking between k_0^{bo} and k_1^{bo} is straightforward, as the indirect effect reduces the rationing costs, hence the need for investment. The rest of the proof analyzes how the investment level under the endogenous regime k_1^{bo} behaves compared to the first-best level k_0^*. The core of the proof relies on the variation of k_1^{bo} with respect to p^w which is given by the implicit function theorem: $\frac{\partial k_1^{bo}(p^w)}{\partial p^w} = -\frac{\partial^2 W_1^{bo}}{\partial k^2} / \frac{\partial^2 W_1^{bo}}{\partial k \partial p^w}$. On the other hand, the first-best k_0^* does not depend on p^w, which allows focusing only on $\frac{\partial k_1^*(p^w)}{\partial p^w}$ for the ranking. We previously proved that the expected social welfare is concave, which ensures that $\frac{\partial^2 W_1^{bo}}{\partial k^2} \leq 0$. The cross derivative $\frac{\partial^2 W_1^{bo}}{\partial k \partial p^w}$ is not necessary everywhere negative, so we need additional analysis. When the parameters are such that the price cap binds for some states of the world, there exists an upper bound p^{w+} and a lower bound p^{w-} such that the price cap never binds or always binds in expectations. When the price cap never binds in expectation, $p^w = p^{w+}$, the equilibrium of the different regimes and the different inefficiencies are equal $k_1^*(p^{w+}) = k_0^*(p^{w+}) = k_1^{bo}(p^{w+}) = k_0^{bo}(p^{w+})$, and the equilibrium capacity price is null. Moreover, at the limit of this upper bound, we find that the cross derivative is negative :

$$\frac{\partial^2 W_1^{bo}}{\partial k \partial p^w}|_{p^w = p^{w+}} = -f(t_1^w(k)) \left(1 - \frac{\partial q_1^w(t_1^w, k)}{\partial k}\right) \frac{\partial t_1^w(k)}{\partial p^w} \frac{1}{k} \int_0^k (p(q, t_1^w) - p(k, t_1^w) dq \leq 0$$

As more capacity increases prices and decreases demand: $\frac{\partial q_1^w(t,k)}{\partial k} \leq 0$, and a higher price cap has a positive net effect on the threshold $\frac{\partial t_1^w(k)}{\partial p^w} = \left(1 + \frac{\partial p^c(k)}{\partial p^w}\right) \frac{1}{f(t_1^w(k))} \geq 0$. The expression captures that even at the limit p^{w+}, the marginal effect from reducing rationing cost already dominates the welfare decrease due to the negative price effect. We find similar results for the lower bound $\frac{\partial^2 W_1^{bo}}{\partial k \partial p^w}|_{p^w = p^{w-}} \leq 0$. An analysis of the behavior of the cross derivative confirms that it is always decreasing between the two thresholds. Hence $\frac{\partial k_1^{bo}(p^w)}{\partial p^w} \leq 0$ for any $p^w \in [p^{w-}, p^{w+}]$ and $k_1^{bo}(p^{w+}) = k_0^*$, which proves the uniform ranking.

We conclude the comparison by studying the ranking between the welfare at the different investment equilibria. Similarly, we have proven that the exogenous regime provides the highest welfare with only a price cap inefficiency and that inefficient rationing decreases welfare: $W_0(k_0^*) \geq W_0^{bo}(k_0^{bo})$ and $W_1(k_1^*) \geq W_1^{bo}(k_1^{bo})$. We are left to study the difference between the exogenous and endogenous regimes at the equilibrium level, that is, $W_0^{bo}(k_0^{bo})$ and $W_1^{bo}(k_1^{bo})$. While not a priori straightforward, we find that there is also a unique ranking between the expected welfare at the equilibrium, which is described in Proposition 4.

Proposition 4. Under the model specifications, there is a unique ranking between the welfare equilibrium such that: $W_0(k_0^{bo}) \leq W_1(k_1^{bo}) \leq W_1(k_1^*) \leq W_0(k_0^*)$

Proof. See Appendix □

The proof focuses on the comparison between $W_0^{bo}(k_0^{bo})$ and $W_1^{bo}(k_1^{bo})$ and relies on three observations : (i), from Proposition 3, for a given p^w, we have $k_1^{bo} \leq k_0^{bo}$; (ii) the two functions are increasing in p^w, and (iii) ΔW_1^{bo} is decreasing and concave in p^w. We next discuss the intuitions behind the results.

Figure 6 illustrates the results. The solid black line gives the first-best welfare function, corresponding to the exogenous equilibrium with only a price cap inefficiency. The black diamond represents the first-best investment level k_0^*. Involuntary rationing is added to represent the public-good nature of the investments. The blue curves represent the new expected social welfare with inefficient rationing under an exogenous regime. The red dashed curve encompasses the effects

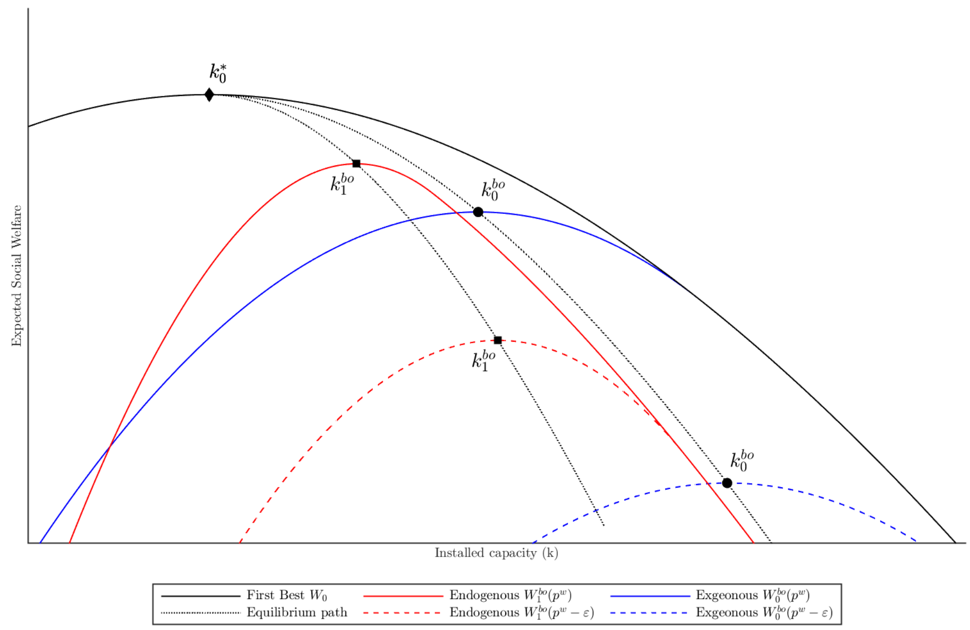


Figure 6: Expected social welfare given different capacity market designs

of the endogenous regime with inefficient rationing. [34] We then vary the price cap level, with a higher value for solid curves compared to the dashed curves. The equilibrium values, which are also the maximum expected welfare reachable within the two market designs, are represented by black squares and circles for the endogenous and exogenous regimes. We also represented the path for each equilibrium for continuous values of the price cap with the dotted line. As expected, following the results in the different Propositions, for a given price cap, the level of investment is always lower under an endogenous regime. The values of the expected welfare at the black squares are always higher than those at the black circles. Hence, within our framework, the endogenous regime always provides higher welfare than the exogenous regime under inefficient rationing.

[34]Note the convergence for the different curves to the right. Above a specific value of k, the price cap never binds in expectation, and inefficient rationing ceases to exist. It only remains the negative price effect of the endogenous regime for the blue curves.

The rationale behind the results lies in two distinct causes : (1) the change of welfare due to the indirect effect as described in $\Delta W_1^{bo}(k)$, for a given investment level, it reduces rationing costs, and leads to negative price effect. (2) the change of investment equilibrium, with previous Proposition 3 stating that $k_1^{bo} \leq k_0^{bo}$. The change in investment level between the two market designs can then be decomposed into two components : (i) the relative comparison between the negative price effect and the reduced rationing costs, and (ii) the relative investment costs between the two market design equilibria. For the second component, the effect is always positive for the endogenous regime, as a lower investment level always implies lower investment costs. The dominance ranking between the opposite effects is ambiguous for the first component. The endogenous reduced rationing costs for a given equilibrium investment level do not always overcome the negative price effect it generates. We even find that having a lower investment level can penalize the endogenous regime regarding the net effect. However, as shown in Proposition 4, saving due to lower investment costs always implies higher welfare under the endogenous regime at the equilibrium.

Figure 7 illustrates the interaction between those opposite effects. The dashed curves represent them under a unique investment level (as described by $\Delta w_1^{bo}(k)$), which is arbitrarily taken equal to the equilibrium under the exogenous market design k_0^{bo}. As expected, the difference in investment costs (black curves) is null. For low values of the price cap, the reduced rationing costs dominate the negative price effect. As the price cap increases, there is a switch, and the negative price effect dominates. We then show the actual comparison between the two regimes at their respective equilibrium with the solid curves. In that case, the gain in avoided rationing costs decreases, and the negative price effect increases. However, the lower investment level saves investment costs, represented by the solid black curves. In that case, the sum of the black and blue curves always dominates the red curve, which implies higher welfare under the endogenous regime at the equilibrium.

We end this analysis by focusing on comparing the first-best solution under the exogenous regime and the first-best under the endogenous regime in terms of expected social welfare. While there can be a clear-cut answer on the ranking between the first best in terms of investment level, this does not transfer in a similar fashion in terms of expected social welfare. In other words, we state

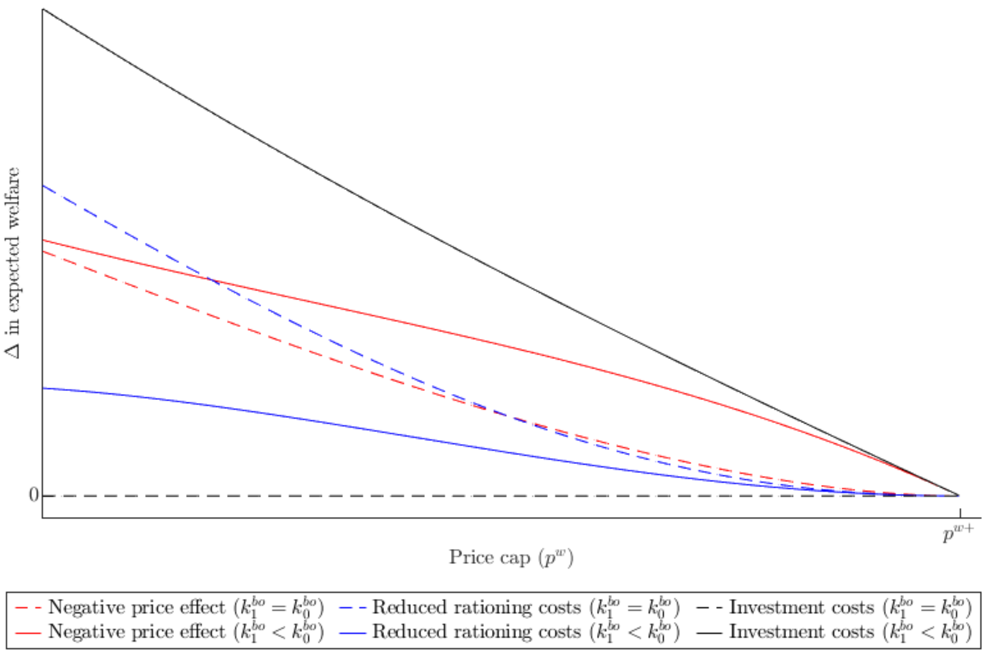


Figure 7: Decomposition of the change in welfare at the equilibrium between endogenous and exogenous market design with inefficient rationing

that the endogenous regime can potentially generate additional benefits compared to the exogenous case. When the regulator implements such a regime for the capacity market and adapts its new investment objective, then at the new investment level, the expected social welfare could be higher than what would have been reachable under the exogenous regime. The comparison between the expected social welfare and its net effect depends (i) on the total investment cost, (ii) on the size of the negative effect previously described of the capacity price $\Delta W_1(k, p^c)$, (iii) the gain/loss in terms of avoided rationing cost $\Delta M_1(k)$. The first effect is always positive in favor of the endogenous regime when the result of the Proposition 4 holds, that is, $k_1^* \leq k_0^*$. The second effect is always negative as shown with the Proposition 2. Finally, the third effect can be ambiguous, as shown in the following analysis. To illustrate the conditions under which one effect dominates the other, we develop the model under a new specification based on a discrete distribution. Note that all the previous results holds under this new specification.

The technical implementation and how the different first-best investment levels are found for the exogenous and endogenous regimes are detailed in the Appendix. The central result is that it exists for specific values of the price cap p^w and probability value θ in some cases where the expected social welfare at the first-best investment under the endogenous regime level is either higher or lower than the exogenous regime. Therefore, the endogenous regime is not always beneficial when considering inefficient rationing. We formally describe the existence of the negative effect in the endogenous regime in the following proposition.

Proposition 5. There can exist some functions $p_1^w(\theta)$ and $p_2^w(\theta)$ such that $\forall \theta \in [p_1^w(\theta), p_2^w(\theta)]$ we have $W_1^{bo}(k_1^*) \leq W_0^{bo}(k_0^*)$. Outside the boundaries we always have $W_1^{bo}(k_1^*) \geq W_0^{bo}(k_0^*)$.

The proof of such negative effect cases is based on the type of first-best emerging in the model for the exogenous and endogenous cases for given values of $\{\theta, p^w\}$. For sufficient low value of the investment cost $r \leq \tilde{r}$ and of the price cap $p^w \leq \tilde{p^w}$, there exist three different first-best for the exogenous regime : $k_0^* = \{q_0^w, k_0^{*,1}, k_0^{*,2}\}$, and three for the endogenous regime $k_1^* = \{q_1^w, k_1^*, q_1\}$. The conditions on r are sufficient so that we always have only six distinct comparison cases: $\{q_0^w, q_1^w\}$,$\{k_0^{*,1}, q_1^w\}$,$\{k_0^{*,1}, k_1^*\}$,$\{k_0^{*,2}, k_1^*\}$,$\{k_0^{*,2}, q_1\}$. We find that for the first two cases, the expected social welfare at first best is always superior under the endogenous case. For the other case, the conditions for having a negative difference in a specific comparison case can be contained within the conditions of existence of the comparison case.

A deeper study of the different components of the difference between the expected social welfare shows that the window during which the negative difference occurs is mainly due to the fact that the absolute $M(k)$ can be decreasing with respect to the probability value θ. Therefore, the function $p_1^w(\theta)$ denotes the limit at which the gains in terms of avoided welfare loss under the exogenous regime are higher than the gains of investment cost and rationing cost under the endogenous regime (at $\{k_0^{*,1}, q_1^w\}$). Finally, for the last comparison case, the avoided welfare loss decreases with respect to θ, which is one of the reasons why for θ above $p_2^w(\theta)$, the net effect becomes positive again. Note that the positive difference is only due to the avoid investment cost in the last comparison case (at $\{k_0^{*,2}, q_1\}$), the expected rationing cost being lower under the exogenous case. We show in Figure

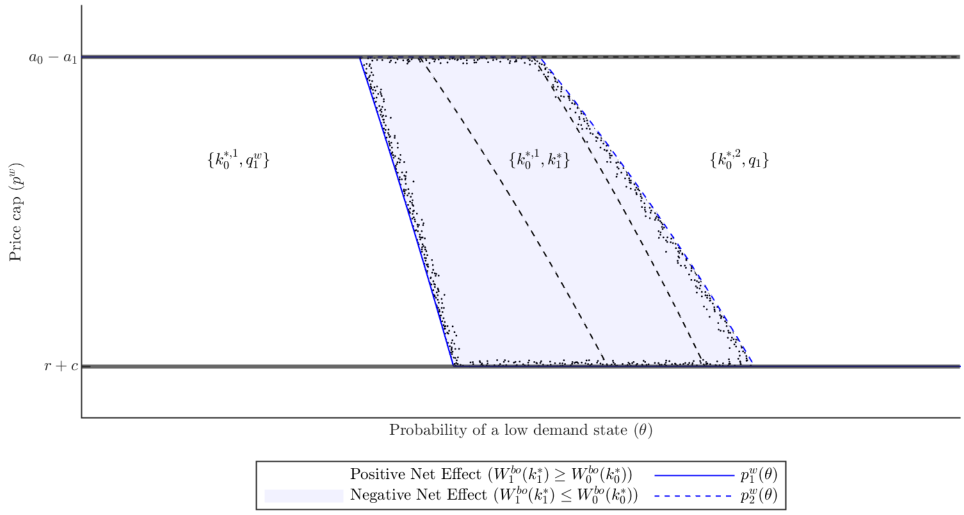


Figure 8: Sign of the delta between expected social welfare at the first-best investment level

8 the relation between the sign of the difference and the two parameters p^w and θ, as well as the boundaries for the first-best (in black dashed lines) and the boundaries for the sign change. The upper and lower limits on p^w originate from the initial assumptions to ensure an equilibrium.

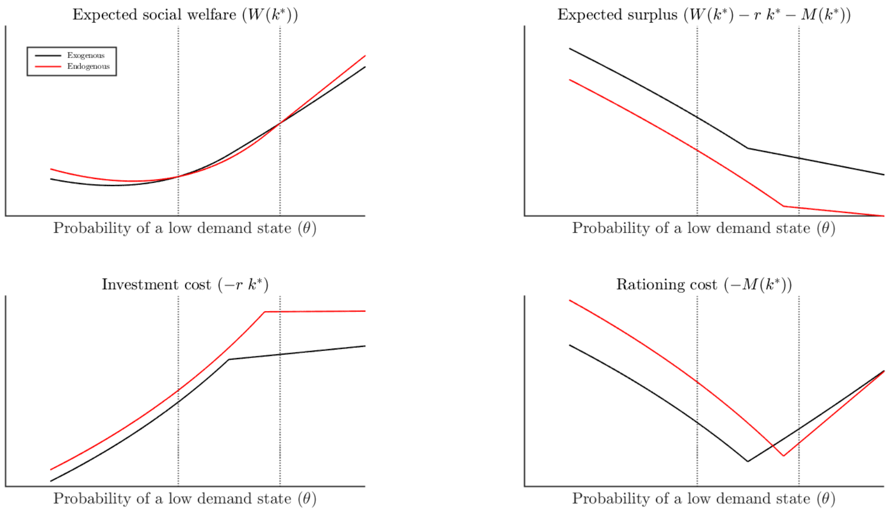


Figure 9: Losses and gains from the exogenous and endogenous regime for a given price cap p^w and different values of θ

5 Capacity demand allocation design

We now extend the previous model to analyze allocations. Under those regimes, the quantity allocated to the retailers (or directly to the consumers) becomes dependent of the demand's current realization. While this issue is not significantly relevant in a context where the inadequacy between capacity installed and the demand level is solely manageable through the supply side, we demonstrate in this section that indirect incentives created by considering the realized demand in the market design can provide additional benefits to the system. This is particularly true with the presence of inefficient rationing. We apply the initial model to two settings : (i) when the allocation depends on the realized market share of retailers playing '*A la Cournot*' (ii) when the capacity market is fully decentralized where retailers by themselves the capacity.

5.1 Retailers Market share allocation

Under this market design, the capacity allocation depends on the retailers' realized quantity sold to the final consumers. To represent retailers' market share, we assume in this section that they play *à la Cournot.* This allows us to indirectly shed light on the effect of imperfect competition in the retail market, which is an overlooked subject in the literature, both in terms of investment decisions in capacity and in the design of capacity market. We underline that having different degrees of competition in the retail market has a direct effect on the capacity cost allocation sustained by final consumers. Still, it also indirectly impacts the efficiency of the system in a somewhat different way than imperfect competition on the supply side.[35]

5.1.1 Market equilibrium with Cournot and without a capacity market

We start the analysis by studying the impact of having retailers playing *à la Cournot* in the retail market without any capacity market. The technical details and equations are described in the Appendix section for clarity. We do not include a price cap in this first analysis as it would not change the results. We denote $mp(q)$ the markup associated with the market power in the retail market such that $mp(q) = -\frac{q}{n}p_q(q)$[36]. Similarly to the other cases, we denote $k^*_{0,n}$ the investment level that maximize the expected social welfare $W_{0,n}(k)$ under imperfect competition such that $k^*_{0,n} = \{k : \phi_{0,n}(k) = r\}$, with :

$$\phi_{0,n}(k) = \int_{t_{0,n}(k)}^{+\infty} (p(k,t) - c)dF(t)$$

With $t_{0,n}(k)$ the threshold value such that this is the first state of the world for which the wholesale price adjusted by retailers' market power equals the marginal cost: $t_{0,n} = \{t : p(k,t) - mp(k) = c\}$. The initial general assumptions ensure that the expected social welfare is concave under the *Cournot* competition. The market equilibrium, such as the expected marginal revenue, equals the marginal cost of providing an additional investment: $k^n_0 = \{k : \phi^n_0(k) = r\}$ with :[37]

$$\phi^n_0(k) = \int_{t_{0,n}(k)}^{+\infty} (p(k,t) - mp(k) - c)dF(t)$$

[35] See, for instance, (Léautier, 2016) or Zöttl (2011) for an analysis of market power on the supply side.

[36] Note that $mp(q) > 0$ as $p_q(q) < 0$, we simplify $\frac{\partial mp(q)}{\partial q} = mp_q(q)$ and $\frac{\partial mp(q)}{\partial n} = mp_n(q)$.

[37] Recall that under perfect competition, the relation between the wholesale price and retail price is given by $p^s(q,t) = p(q,t) + \frac{q}{n}p_q(q)$

The following Lemma sums up the results: we find that market power in the retail market lowers the investment level beyond the market power's direct effect. The market investment level is different from the optimal investment level even when maximizing the welfare function given the market power in the retail market.[38]

Lemma 7. For every $n \in [2, \infty[$, imperfect competition in the retail market leads (i) to a lower first best capacity level compared with the optimal investment level $k_0^* \geq k_{0,n}^*$ (ii) to a lower market equilibrium in terms of investment level $k_{0,n}^* \geq k_0^n$. The optimal capacity payment z_n is equal to the expected markup of retailers in the retail market during peak periods:

$$z_n(k) = \int_{t_{0,n}(k)}^{+\infty} mp(k) dF(t) \tag{10}$$

Proof. See Appendix □

With imperfect competition, the first best investment level is lower due to a shift in the occurrence of on-peak periods. It can be showed by comparing the expression $\phi_0(k)$ and $\phi_{0,n}(k)$ due to the presence of the threshold value $t_{0,n}(k)$. Similarly, the inefficient market equilibrium is also lower due both to the threshold and to a lower expected price $p(q,t) - mp(q)$ as shown in the expression of $\phi_0^n(k)$.

We illustrate this result with the numerical illustration used in example 1. In figure 10, we compute the expected social welfare for different values of capacity, and we show the cumulative effect of imperfect competition in the system. The black curve represents the case with perfectly competitive retailers. The black point 1 is the first best investment level k_0^* that maximizes this surplus. We use a price cap to represent the first inefficiency that implies a low market equilibrium represented by the black square 3. The results in Lemma 7 demonstrates that with imperfect competition in the retail market : (i) the expected social welfare is lower, which is represented by the red curves and implies a lower first-best investment level at the red point 2, (ii) the market equilibrium is also

[38] This result has important regulatory implications. Indeed, we state that the welfare-maximizing investment level, given the imperfect competition in the retail market, is different from the welfare-maximizing investment level in a perfectly competitive market. Therefore, reaching a competitive investment level could cause significant harm, potentially greater than the welfare loss generated by the inefficient market equilibrium.

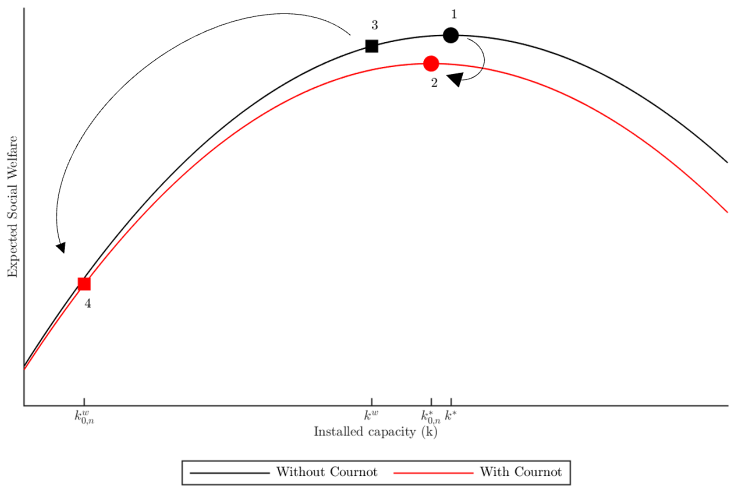


Figure 10: Expected social welfare, first-best and market investment level with and without imperfect competition in the retail market and with a price cap in the wholesale market

lower compared to the initial case with only a price cap. The red square 4 encompasses the two inefficiencies and is significantly different from the initial market equilibrium.[39]

5.1.2 Market equilibrium with Cournot and with a capacity market

We now introduce the capacity market by describing the new market design regime as follows.

Assumption 4. A single entity builds a demand function in the capacity market and buys a level of capacity k for a price $p^c(k)$ given the supply function described in 3. Then it allocates the full cost $kp^c(k)$ directly to the retailers. The share of the capacity cost is based on their realized market share in the retail market. For a retailer i this share is defined as $\frac{q_i}{q_i+q_{-i}}$ with q_i its quantity sold on the retail market and q_{-i} the quantity sold by its competitors.

[39]The value k_0^n is found by estimating the expected marginal revenue of producers $\phi^w_{0,n}(k)$, which is lowered both by the price cap and retailers' market power.

The first implication of ex-post allocation concerns the last stage, which is when the retail market clears. We rewrite the retailers' profit function by including an endogenous ratio in the retailer profit function, as shown in the following equation.

$$\pi_i^r(q_i, k) = q_i(p(q) - p^s) - p^c(k)k\frac{q_i}{q_i + q_{-i}}$$

Contrary to the previous section, we do not need to assume any tariff hypothesis for the capacity cost allocation as it directly affects retailers' profit at the margin. We focus our analysis on symmetric equilibrium. We drop the notation with t as the state of the world is known at this stage. With $q_{-i} = \sum_{j \neq i}^{n} q_j$. We find the best-response function of a retailer i with the first-order conditions :

$$BR_i(q_j) = \max_{q_i} \pi_i^r(q_i, k) \Longleftrightarrow p(q) + q_i p_q(q) - p^s - p^c(k)k\frac{q_j}{(q_i + q_j)^2} = 0$$

The main results for the existence of equilibrium are stated in the following Lemma:

Lemma 8. At the symmetric equilibrium, retailers's profit function is concave if the capacity cost is not too important, that is if the following condition holds :

$$q^2\left(\frac{n+1}{n-1}p_q(q) + \frac{q}{n-1}p_{qq}(q)\right) > p^c k$$

The quantities are strategic substitutes whenever the following condition holds :

$$kp^c(k)\left(\frac{n-2}{n}\right)\frac{1}{q^2} \geq p_q(q) + \frac{q}{n}p_{qq}(q)$$

Proof. See Appendix □

Note that when $n = 2$, the equilibrium is always unique and stable. This observation comes from the classical decreasing marginal returns of the Cournot literature Vives (1999). When $n > 2$, the lemma states that the capacity market allocation design induces a stricter condition on the marginal returns, which needs to consider the additional cost in the retailer's profit. Using the first-order conditions and the symmetry between the retailers, the Cournot equilibrium in the retail markets allows to define the endogenous retailer demand function in the wholesale market:

$$p_n(q) = p(q) - mp(q) - p^c(k)k\frac{1}{q}\frac{n-1}{n}$$

The concavity condition stated in Lemma 8 is the same as assuming that the demand function in the wholesale market is decreasing. The equilibrium in this market design is similar to the endogenous regime in the previous section due to the effect of the capacity price on the final demand. Therefore, we can again define the new periodic thresholds between on-peak/off-peak/binding price cap periods. We denote them $t_n(k)$ and $t_n^w(k)$ such that the expected value of $p_n(q)$ is equal to respectively the marginal cost and the price cap, that is $t_n(k) = \{t : p_n(q,t) = c\}$ and $t_n^w(k) = \{t : p_n(q,t) = p^w\}$. We denote the corresponding quantity $q_n(t)$ and $q_n^w(t)$ such that $q_n(t) = \{q : p_n(q,t) = c\}$ and $q_n^w(t) = \{q : p_n(q,t) = p^w\}$.

The indirect effect of this market design on the system can be shown in the following expression of the supply function in the capacity market:[40]

$$p^c(k) = r - \left(\int_{t_n(k)}^{t_n^w(k)} (p(k) - mp(k) - c - p^c(k)\frac{n-1}{n})dF(t) + \int_{t_n^w(k)}^{+\infty} (p^w - c)dF(t) \right)$$

It is composed of the marginal opportunity cost of providing an additional capacity to the system. Hence, the marginal cost of installing a capacity r and the expected marginal rent received during on-peak periods are defined in the parenthesis. This second expression shows that when the capacity price does not bind, the producers receive a lower share of the demand function of the final consumers due to the effect of retailers' market power and the depressing effect of the market design. The endogeneity created by the market design can be found in the fact that the expression $p^c(k)$ appears in both terms via the demand function in the wholesale market and via the threshold

[40]For clarity, we do not expose the full effects of this market design on the expected social welfare nor on the first-best solution. Indeed, it has consequences similar to the endogenous design; therefore, it implies a depreciation of the demand for the final good. It changes the expected surplus during off-peak states of the world and the occurrence between offpeak / on-peak periods, and the first-best investment level is lower than an allocation without indirect effect. The full technical details are exposed in the Appendix.

value $t_n(k)$ and $t_n^w(k)$. Note the difference with the requirement in the first part of the integrals, where the capacity cost adders are dependent on n. Another difference lies in the results of Lemma 8. Indeed, it states that the demand function on the wholesale market is not fully defined for every pair of capacity price/investment level. This is particularly true for high capacity prices and high investment levels in conjunction with lower demand for the final good. In this case, the capacity cost is too high, implying an absence of trade in the wholesale market. It has some implications for the estimation of optimal outcomes as for high investment levels, the expected social welfare function is not always concave, as opposed to the previous analysis. [41]

The following Proposition summarizes the main effect of an ex-post allocation based on the realized market share. Allocating the capacity market cost based on retailers' realized market share provides an intermediate indirect effect between an exogenous regime price and an endogenous regime. We extend the previous section to take into account imperfect competition. We denote $k_{1,n}^*$ the first-best under an endogenous design with imperfect competition similarly. That is, the capacity price is allocated on a variable basis directly to the final consumers. In that case the marginal expected social welfare such as $k_{1,n}^* = \{k : \phi_{1,n}(k) = r\}$ is given by

$$\phi_{1,n}(k) = \int_0^{t_{1,n}(k)} \frac{\partial q_{1,n}(t)}{\partial k} p^c(k) dF(t) + \int_{t_{1,n}(k)}^{+\infty} (p(k,t) - c) dF(t)$$

As with $k_{0,n}^*$, we simply extend the previous analysis to take into account imperfect competition.

Proposition 6. (i) The new first-best solution in terms of investment level under the market share allocation regime exists if the following condition holds $\frac{\partial p^c(k)}{\partial k}(p_{qq}(k) - mp_{qq}(k)) \geq \frac{\partial^2 p^c(k)}{\partial k^2}(p_q(k) - mp_q(k))$.

(ii) If it exists it solves $k_n^* = \{k : \phi_n(k) = r\}$, with $\phi_n(k)$ defined as follow

[41] One way to see it is by defining a state of the world as the threshold between the case where the maximum price on the wholesale market is null at the quantity exchanged during off-peak periods. Namely we define the threshold $t_n^-(k)$ such as the concavity condition in Lemma 8 is strictly null, that is :$t_n^-(k) = \{t : q_n(t)^2 \left(\frac{n+1}{n-1} p_q(q_n(t)) + \frac{q_n(t)}{n-1} p_{qq}(q_n(t))\right) - p^c k = 0\}$. Whenever this threshold exists, then the expected social welfare is defined as follow:$\int_{t_n^-(k)}^{t_n} \int_0^{q_n(t)} (p(q,t) - c) dq dF(t)$. At the second derivative of the welfare function, the threshold value plays a role and can lead to the non-concavity of the expected welfare.

$$\phi_n(k) = \int_0^{t_n(k)} \frac{\partial q_n(t)}{\partial k} p^c(k) \frac{n-1}{n} dF(t) + \int_{t_n(k)}^{\infty} (p(k,t) - c) dF(t) \tag{11}$$

(iii) The first-best investment level is lower than the first-best under exogenous design and higher than the first-best under the endogenous regime ($k^*_{1,n} \leq k^*_n < k^*_{0,n}$). Moreover, the reverse is true for the expected social welfare at the first-best investment level; the welfare is higher under an exogenous regime but lower under an endogenous regime compared to the market share allocation.

Proof. See Appendix □

The capacity cost adder when $n = 2$ is equal to half of the cost adder of equation 7 increases with n. When $n \to +\infty$, the capacity cost is entirely allocated to the consumer, mimicking the exogenous equilibrium. This Proposition states that increasing competition in the retail market increases the burden of consumers' capacity prices. Hence, the negative effect observed in the regime with endogenous capacity prices is now shared between retailers and consumers. By extension, we have the same results for the endogenous regime when we take into account inefficient rationing with the ex-post market share allocation. Namely, the depressing effect shown in Proposition 6 will both lower the expected surplus and the expected rationing cost, hence having an ambiguous effect on the optimal outcome.

We illustrate our results using the numerical example described in 1. In figure 11, we show the expected social welfare under various assumptions. The black curve represents welfare when there is no capacity market (or an exogenous capacity market), and the retail market is perfectly competitive. Then, the blue solid curve stands for the new welfare function when we assume the retailers are playing *a la Cournot*. Note that the optimal payment defined in equation 10 in Lemma 7 is not the payment necessary to reach point 2 to point 1. Instead, the optimal payment is necessary to reach the market equilibrium represented by the dashed vertical line and point 2. In other words, increasing the level of investment does not change the level of competition in the retail market and does not allow an increase in the expected social welfare beyond the blue curve. In the same figure 11, we also illustrate the results of Proposition 6. The dashed blue curve represents the expected social welfare given the effect of the capacity market design based on realized market

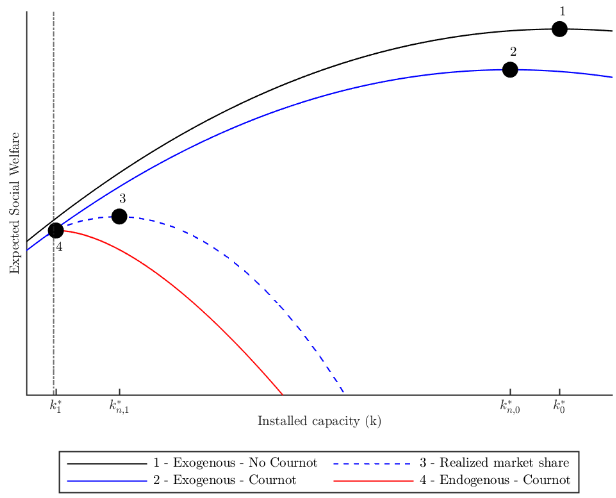


Figure 11: Expected social welfare under the market share allocation ($n = 5$)

share. Similar to the endogenous case, it depreciates the social, such as an increase in the level of investment, increasing the difference between the two welfare. The new first-best investment level noted by point 3 is now lower than the first-best of the initial stat note by point 2. However, compared to the endogenous case of the previous section, the figure shows that this first-best level is always higher than the previous one noted by point 4, both in terms of investment level and in terms of expected social welfare.

5.1.3 Extension - capacity market, inefficient rationing, and retail market structure

We now extend our analysis to the relation between the degree of competition in the retail market and the outcomes of the paper. In this extension and using our analytical framework, we show that a change in market structure can have ambiguous effects both in the capacity market outcome and with respect to the determination of the optimal capacity level. This aspect of essential goods has been relatively less studied than the supply side. Therefore, in this extension, we describe the effect of different market structures on the optimal and market investment level with and without inefficient rationing.

The first level of the analysis is to consider the initial framework without a capacity market and with only missing money inefficiency caused by a price cap. In this case, Lemma 7 states that an increase in the number of retailers always increases both the investment and expected welfare at the first-best level. This result is in agreement with the literature on market power. In this section, we focus both on the change in optimal investment level and on expected social welfare at the same optimal investment level. The derivative of the expected social welfare in this case can be written as follows:

$$\frac{\partial W_{0,n}(k)}{\partial n} = \int_0^{t_n(k)} \frac{\partial q_{0,n}(t)}{\partial n} mp(q_{0,n}) dF(t)$$

With $\frac{\partial q_{0,n}(t)}{\partial n} = \frac{mp_n(q_{0,n})}{p_q(q_{0,n}) - mp_q(q_{0,n})}$. The derivative is positive as the markup decreases with n but increases with q. It proves that $\frac{\partial W_{0,n}(k)}{\partial n} > 0$. The implementation of a capacity market under an exogenous regime does not modify the result. Increasing the number of retailers lowers the market up they impose in the retail market, which increases the demand in the wholesale market. Ultimately, this lowers the cost of providing additional capacity. The effect on the capacity price is ambiguous, as lower market power increases the optimal quantity of investment but lowers the cost of procurement. We turn now to the analysis of the effect of the market structure on the model when there is inefficient rationing. Under this assumption, we find that an increase in competition in the retail market can have an ambiguous effect on the expected social welfare at the optimum. To see this, we express the derivative of the social welfare with respect to n :

$$\frac{\partial W^{bo}_{0,n}(k)}{\partial n} = \frac{\partial W_{0,n}(k)}{\partial n} - \frac{\partial M_{0,n}(k)}{\partial n} = \frac{\partial W_{0,n}(k)}{\partial n} - \int_{t^w_{0,n}(k)}^{+\infty} \frac{\partial J(\Delta_{0,n}k)}{\partial n} + \frac{J(\Delta_{0,n}k)}{\partial \Delta_{0,n}k} \frac{\partial \Delta_{0,n}k}{\partial n} dF(t)$$

With $M_{0,n}(k)$, the cost of inefficient rationing when there is imperfect competition in the retail market but no indirect effect of the capacity market. The net marginal effect on the expected social welfare depends on the representation of the cost of inefficient rationing. Intuitively, an increase in competition increases the demand for the good. Hence, we should expect a higher Δk for a given level of investment k. Therefore, the second derivative in the rationing cost part is positive. The direct effect of n on $J(.)$ is less clear. Let's assume that the cost of inefficient rationing is based on the surplus of the consumers at the investment level k as in our model specification in example **??**. In that case, it is independent of n, and the derivative is null. We show this effect in the following equation with the model specification :

$$\frac{\partial W^{bo}_{0,n}(k)}{\partial n} = \int_0^{t_n(k)} \frac{\partial q_{0,n}(t)}{\partial n} mp(q_{0,n}) dF(t) - \int_{t^w_{0,n}(k)}^{+\infty} \frac{\partial q^w_{0,n}(t)}{\partial n} \frac{1}{k} \int_0^k p(q,t) - p^w dq dF(t)$$

With $\frac{\partial q^w_{0,n}(t)}{\partial n} = \frac{mp_n(q^w_{0,n})}{p_q(q^w_{0,n}) - mp_q(q^w_{0,n})}$, which is also positive. On the other hand, we could have assumed a cost based on the consumer surplus at the quantity exchanged at the price cap, namely $q^w_{0,n}(k)$. But in this case, an increase of n also increases this value via the increase in demand. Hence, the first derivative is also positive. In both cases, the second part of the equation is always negative, counterbalancing the positive effect of $\frac{\partial W_{0,n}(k)}{\partial n}$. This has important policy implications, as we have shown that a lower degree of competition does not always translate into a higher expected social welfare when there is inefficient rationing.

At first sight, a similar observation could have been made when we consider either an endogenous market design or a market share allocation for a capacity market. In that case, an increase in the number of retailers has two effects of opposite sign : (i) an increase in social welfare, which is the common effect of higher competition in a canonical model *à la Cournot* highlighted in the previous equation (ii) a decrease in the social welfare due to the lowering of the consumption associated with a higher capacity cost allocated to the consumers. We provide the marginal value of expected social welfare in the following equation.

$$\frac{\partial W_n(k)}{\partial n} = \int_0^{t_n(k)} \frac{\partial q_n(t)}{\partial n} \left(mp(q_n) + p^c(k) \frac{k}{q^w_n(t)} \frac{n-1}{n} \right) dF(t)$$

As in the previous equation, the expressions in the brackets are positive. The difference between the no capacity market (or exogenous) regime compared to the realized market share (or endogenous) lies in the value of $\frac{\partial q_n(t)}{\partial n}$, which is shown in the following equation.

$$\frac{\partial q_{n,1}(t)}{\partial n} = \frac{\left(mp_n(q_n) - p^c(k)\frac{1}{n^2}\frac{k}{q_n(t)}\right) + \frac{n-1}{n}\frac{\partial p^c(k)}{\partial n}\frac{k}{q_n(t)}}{p_q(q_{n,1}) - mp_q(q_{n,1}) - \frac{p^c(k)}{q_n(t)^2}\frac{n-1}{n}}$$

The denominator of the derivative is always negative, so its sign is always determined by the net effect at the numerator. As previously stated, an increase in n lowers the cost of providing an additional capacity and increases the optimal investment level that should be bought in the capacity market. It creates the potential ambiguity in the value of $\frac{\partial p^c(k)}{\partial n}$. However, at the optimal level, the envelop theorem ensures that only the direct effect of n on the supply curve in the capacity market should be considered, therefore excluding the indirect effect on the optimal investment level. Therefore : $\frac{\partial p^c(k)}{\partial n} < 0$. All in all, we have $\frac{\partial q_{n,1}(t)}{\partial n} > 0$ and $\frac{\partial W_n(k)}{\partial n} > 0$. The results of this extension are summarized in the following proposition.

Proposition 7. For every capacity market regime without inefficient rationing, the expected social welfare at the first-best investment level is increasing with n.

However, when including inefficient rationing, for every capacity market regime, the increase of competition always increases the optimal investment level, but it has an ambiguous effect on the expected optimal social welfare. When the marginal gain of expected surplus during off-peak periods is higher than the marginal loss of inefficient rationing, then the expected social welfare increases.

We illustrate those results using the model specification of example 1. Figure 12 shows how the degree of competition in the retail market, the market design, and the assumption with respect to the inefficiencies can significantly impact the expected social welfare at the first-best investment level. While the latter is always increasing in n, this is not always the case for welfare. As expected, in the canonical market design with no inefficient rationing (black solid curve), the welfare strictly increases in n. When we implement the endogenous regime - or the market share allocation

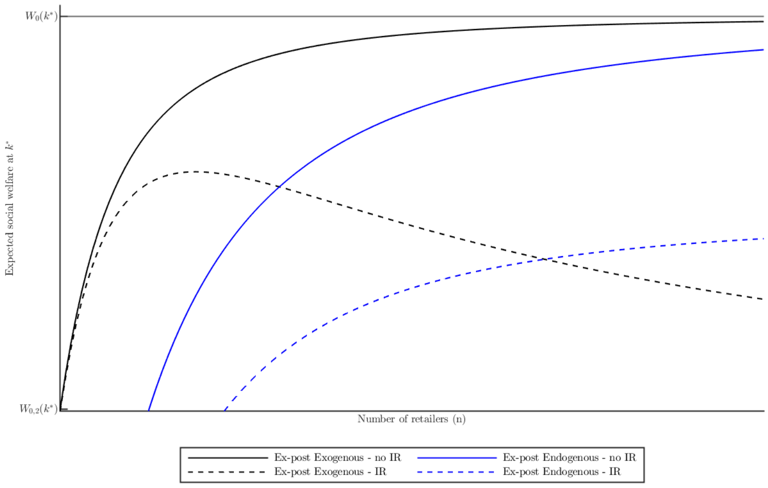


Figure 12: Expected social welfare at the first-best investment level for different levels of competition in the retail market and for different market designs

(blue solid curve), the welfare is still increasing, as the negative effect of more investments is not considered at the optimal level. The central result lies in the case when inefficient rationing is represented. In the first market design (dashed black curve), the expected social welfare at first best is a concave function, implying that an additional retailer provides more harm than benefits to the system above a certain degree of competition in the retail market. In other words, the marginal cost associated with higher demand and a higher cost of inefficient rationing is greater than the gain in lower market power. Finally, the figure shows another benefit of either an endogenous market design or a market share allocation with inefficient rationing (dashed blue curve). Indeed, the expected social welfare becomes an increasing function. It means that the depreciating effect of the capacity price on the demand is not symmetric between the marginal gain and loss of an increase of n. To say it differently, it lowers more the loss associated with the increase of the rationing cost than it decreases the expected surplus during off-peak periods, as shown in the previous section.

5.2 Retailers individual allocation

This last section provides the first analysis of a fully decentralized capacity market. This market design regime takes the furthest step towards accounting for the final electricity demand in the capacity market allocation. Each retailer must purchase their capacities in the capacity market. An entity only monitors ex-post the level of capacities and compares it to each retailer's sales. A penalty mechanism is implemented if there is any difference between the two quantities. We formally describe this market design regime as follows:

Assumption 5. A regulated entity mandates the retailers to buy the capacity on the capacity market given their realized sales in the retail market for a price $p^c(k)$ given the supply function described in 3. For each retailer, if their individual realized sale quantity on the retail market q_i is above their individual purchase quantity on the capacity market k_i, the regulated entity imposes a unitary penalty S, such that the penalty mechanism total cost for a retailer is $S(q_i - k_i)$.

One of the critical features of this regime concerns the case when a retailer is in negative deviation, i.e., has sold more on the retail market than he has bought capacity in the capacity market. In this case, he suffers a penalty, which results in a payment from the retailer to the regulated entity by a unitary amount of S, with $S \geq 0$ being an administratively fixed value.[42]

We process as follows to describe the implication of this market design regime: (i) We analyze the case when the penalty value is null, allowing us to define retailers' fundamental behavior in the capacity market. We show that it significantly depends on the existence of a price cap in the wholesale market and on the level of imperfect competition in the retail market. We develop our results on the idea that retailers act similarly as producers when given the opportunity to participate in the decentralized capacity market. With imperfect competition and a price cap, their expected profit in the retail market depends on the level of capacity. It means that we can also define an expected unitary rent, similarly to the producers' rent $\phi(k)$. Moreover, the existence of a capacity market implies that they can also choose to invest, with the capacity price acting as an investment cost. (ii) Then, we introduce the effect of the penalty on retailers' behavior,

[42]Some remuneration mechanisms can exist so as to reward retailers who have provided additional capacity, but as we focus on symmetric equilibrium, they do not play a role in the outcome.

punishing retailers for having too little capacity. We compare the market equilibrium between the two previous cases and with the other market designs. (iii) Finally, we analyze the implications in terms of expected social welfare and the first best investment level of the decentralized capacity market. This is particularly relevant in inefficient rationing that creates an additional cost for social welfare due to insufficient capacity. We first provide a definition and a comparison of the first best investment level implied by a decentralized capacity market, and then we analyze the market equilibrium. Note that while the outcome relies on a decentralized supply and demand side, the existence of the penalty mechanism acts as a tool for the regulated entity to incentivize to a particular market equilibrium.

5.2.1 Equilibrium in a decentralized capacity market without a penalty mechanism

A fully decentralized capacity market gives retailers an opportunity to choose the level of investment. Therefore, we show that the existence of a decentralized capacity market can provide additional investments even though there is no penalty system implemented. To do so, notice the similarity between producers and retailers, in the sense that at the margin, retailers can have a positive value for increasing the level of investment. We provide in the following equation the sum of expected profit for n retailers when there is imperfect competition in the retail market and when there is a price cap in the wholesale market:[43]

$$\Pi_d^w(k) = \int_0^{t_{0,n}(k)} q_{0,n}(t)mp(q)dF(t) + \int_{t_{0,n}(k)}^{t_{0,n}^w(k)} kmp(k)dF(t) + \int_{t_{0,n}^w(k)}^{\infty} k(p(k,t) - p^w)dF(t)$$

The expected profit is composed of the markup differential between wholesale and retailer prices whenever the price cap does not bind and the expected price differential between the price cap and

[43] For clarity, we assume imperfect competition in the retail in this section. The initial threshold values in this case are similar to the exogenous case, our without capacity market, but we adjust by adding a n to the subscript 0. The extension to the perfectly competitive case is straightforward. For instance, when there is no price cap, the third part of the equation is null.

the retail price when the price cap binds.[44] The marginal value for retailers of a marginal increase of capacity is given in the following equation.

$$\frac{\partial \Pi_d^w(k)}{\partial k} = \int_{t_{0,n}(k)}^{t_{0,n}^w(k)} (mp(k) - kmp_q(k)) + \int_{t_{0,n}^w(k)}^{\infty} (p(k,t) - pw - kp_q(k)) \tag{12}$$

Increasing the level of investment has two effects: (i) It increases the markup during on-peak periods whenever the price cap is not binding. (ii) Retailers can sell an additional unit at a marginal cost equal to the price cap and at the expense of a decrease of the price on the inframarginal quantity whenever the price cap binds. The marginal increase of k does not impact the off-peak periods as the markup is independent of the level of investment. And the marginal effect at the thresholds ($t_{0,n}(k)$ and $t_{0,n}^w(k)$) cancels out. The equilibrium in the decentralized market is found by equating the supply function defined in 3 with the expected marginal value of an additional investment k defined, for instance, in 12 with a price cap. Therefore, similarly to the supply side, we assume that this value acts as a proxy for the demand function in the decentralized capacity market.

Leaving aside the coordination issues between retailers, the two following Lemma summarize the implications of a decentralized market without any penalty system with and without a price cap in the wholesale market. The efficiency of a decentralized capacity market in terms of investment level can be assessed by comparing the value that satisfies the equality with the ones described in the previous sections, such as the optimal condition in equation 1 or the market equilibrium condition in equation 2.

Without a penalty mechanism and without a price cap :

Proposition 8. (i) The market equilibrium in terms of investment level k_0^d is given by the following equality $k_0^d = \{k : \phi_0^d(k) = r\}$ such that

[44]Therefore, we assume that there is no price cap in the retail market. This is a strong assumption, especially since price caps do not necessarily appear explicitly in retail markets. For instance, retailers can be under a form of regulated tariffs that allow some extra profits, or they can have part of their sales covered by forward contracts. However, if a price cap is also implemented in the retail market, it is sufficient to have a positive delta to maintain this effect.

$$\phi_0^d(k) = \int_{t_{0,n}(k)}^{\infty} (p(k,t) + kmp_q(k) - c) dF(t)$$

(ii) In this case, the installed capacity k_0^d, is always above the first-best investment level that maximizes the expected social welfare $k_0^d \geq k_0^*$.

The Lemma stems from the fact that the marginal value for retailers of an additional capacity is always positive whenever there is no price cap in the wholesale market. Therefore, when facing a decentralized capacity market and the supply function described as in the equation 3, there is a tradeoff between gaining a marginal increase in the expected profit and sustaining an additional cost due to the capacity prices. The inequality $k_0^d \geq k_0^*$ follows from the observation that $\phi_0^d(k) \geq \phi_0(k)$. If there is a price cap in the wholesale market, Lemma defines the new equilibrium as well as their ranking.

Proposition 9. (i) There can be two investment equilibriums, where :

(i.a) The first equilibrium k_1^d is given by the following equality $k_1^d = \{k : \phi_1^d(k) = r\}$, with such that :

$$\phi_1^d(k) = \int_{t_{0,n}(k)}^{t_{0,n}^w(k)} (p(k,t) + kmp_q(k) - c) dF(t) + \int_{t_{0,n}^w(k)}^{\infty} \left(p^w - c + mp(k) + k \frac{\partial mp(k)}{\partial k} \right) dF(t)$$

(i.b) The second equilibrium installed k_2^d, if not constrained by $k_{0,n}^w$, is given by the following equality $k_2^d = \{k : \phi_2^d(k) = r\}$, with

$$\phi_2^d(k) = \int_{t_{0,n}(k)}^{t_{0,n}^w(k)} (p(k,t) - kmp_q(k) - c) dF(t) + \int_{t_{0,n}^w(k)}^{\infty} (p(k,t) + kp_q(k) - c) dF(t)$$

(ii) (ii.a) If the price cap is sufficiently binding then the two market equilibrium are an upper and lower boundary of the initial first-best, that is respectively: $k_1^d \geq k_{0,n}^* \geq k_2^d$ (ii.b) If $k_1^d \geq k_2^2$ holds then $k_1^d \geq k_{0,n}^* \geq k_2^d$ also holds. (ii.c) Increasing the degree of competition in the retail market makes the first equilibrium less attractive than the second equilibrium.

The second Lemma is due to the structure of the expected profit whenever there is a price cap in the wholesale market. The threshold effects of the capacity constraint imply that the expected profit can have two local maxima. Whenever the investment level is sufficiently low such that off-peak periods merely exists ($t_{0,n}(k)$ almost null), the expected social welfare is a well-defined concave function, which gives the second equilibrium. However, as soon as the investment level starts increasing, the second part of equation 12 is negative and outweighs the first part. It means that when the price cap binds the marginal loss associated with the marginal cost p^w and the decrease in the inframarginal revenue $kp(k)$ is higher than the price gain $p(k,t)$ and the markup gains when the price cap does not bind (first part of the equation). It implies that the profit function becomes convex. For high investment value, the price cap never binds, which implies an expected profit for retailers as in the no price cap case[45]. Hence, for those investment values, the expected profit is an increasing function of k, which gives the first equilibrium. The existence constraint on the second equilibrium by the initial market equilibrium without a capacity market $k_{0,n}^w$ is straightforward. While retailers might not want this level of investment in the case that $k_{0,n}^w \geq k_2^d$, they cannot force producers to invest less than what they would have done without a capacity market. To say it differently, whenever the inequality holds, this means that the second market equilibrium implies a fully inefficient capacity market with a null capacity price and an unchanged level of investment. The second observation shows that there can be a clear ranking between the equilibrium such that equilibrium is always above the equilibrium, and the initial first-best investment level is comprised between the two (i.e. $\phi_1^d(k) \geq \phi_{0,n}(k) \geq \phi_2^d(k)$). The condition on p^w for having $k_1^d \geq k_0^* \geq k_2^d$ has an interesting economics interpretation. When studying the equilibrium condition for k_0^* and k_1^d, we find that the marginal expected revenue for retailers during on-peak periods when the price cap does not bind (first part of $\phi_1(k)$) is always superior to the marginal expected value of social welfare during the same period (the value of $\phi_{0,n}(k)$ for every $t \in [t_{0,n}(k), t_{0,n}^w(k)]$). On the other hand, the marginal expected revenue for retailers during on-peak periods when the price cap binds

[45]The difference between $\phi_0^d(k)$ and $\phi_1^d(k)$ lies in the value of the supply function that takes into account the price cap.

is either superior or inferior to the social welfare value. depending on the value of p^w. When the price cap is sufficiently binding, the first effect always dominates the second effect, which implies that $k_1^d \geq k_{0,n}$. Similar opposite effects can be found when comparing k_0^* and k_2^d. Finally, the last observation comes from the fact that whenever the marginal expected revenue for retailers during on-peak periods when the price cap binds is higher for $\phi_1^d(k)$ than for $\phi_2^d(k)$ (i.e. the second part of their expression), then we always have a marginal expected revenue for retailers during on-peak periods when the price cap binds higher than the marginal expected value of social welfare during the same period. To illustrate the ranking, our model specification provides a clear-cut answer for the threshold value of p^w: whenever the price cap is lower than the lowest possible wholesale price at the initial first-best investment level $k_{0,n}^*$ such that $p^w = \lim_{t\to\infty} p(k_{0,n}^*, t) - mp_(k_{0,n}^*)$, then the ranking holds. Note that this is a tighter condition than the initial assumption on p^w defined in the original model. Finally, the result with respect to an increase in the number of retailers highlights the relation with the choice of equilibrium even though we do not model coordination. Our result originates from two observations : (i) expected retailers' profit is lower when there is an increase of n for a high level of investment k, and (ii) the second equilibrium decreases less with respect to n compared to the first equilibrium. The economic intuition behind these results is that the first equilibrium and the expected profit with a high level of investment rely mostly on the expected markup. Hence, an increase of n necessarily lowers its profitability. On the other hand, the second equilibrium, and the expected profit with a low level of investment, relies relatively more on the price differential between the retail price $p(k,t)$ and the marginal cost equal to the price cap p^w, both being independent of n.

We illustrate this result using our model specification. We show in figure 13 the relation between the expected retailers' profit, the investment level, the presence of a price cap in the wholesale market, and imperfect competition in the retail market. When there is no price cap (solid lines), the profit function is strictly increasing with k, with a convergence toward a plateau starting at the right square corresponding to profit-maximizing investment level [46]. Given the investment value for retailers and the presence of a capacity market. The intersection between the supply and demand functions in the retail market leads to the right diamonds, the first equilibrium.

[46]Beyond this investment value there is only off-periods which lead to expected profit independent of k

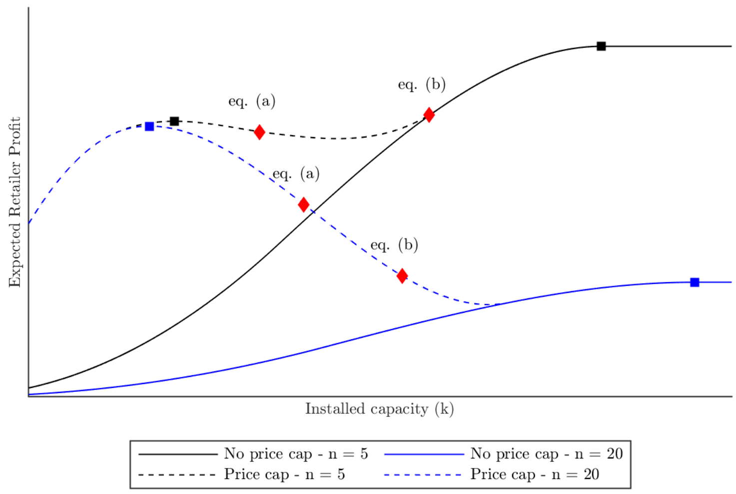


Figure 13: Expected retailers profit with respect to the investment level, the presence of a price cap in the wholesale market, and imperfect competition

This level is similar to assuming that retailers are investing into capacity, such as the marginal revenue is given by equation 12 and the marginal investment cost is given by the supply function of producers in the retail market given by equation 3. As expected, an increase in the degree of competition leads to a decrease in the expected profit, as shown by the difference between the black and blue lines. On the other hand, when introducing a binding price cap, we observe a significant modification of the expected profit function. For a low value of k the profit is concave, and the profit-maximizing level is given by the left squares. An intermediate increase of k leads to a drop in the profit function. Then, the expected profit converges toward the no-price case, as the price case does not bind anymore. While the investment level that maximizes retailers' profit function can be significantly lower than in the no-price cap case, it does not necessarily imply that it is the investment level installed by the producers. It is straightforward to assume that if facing a lower demand of capacity compared to their initial market equilibrium without a capacity market, producers are still going to invest in their market equilibrium level. In this case, the capacity market does not provide any additional capacity. Those investment levels are given by the left diamonds and correspond to the first equilibrium. Finally, the figure shows the ambiguity of implementing a decentralized capacity market. For a low value of n, the expected profit at the second equilibrium is higher than the expected profit at the first equilibrium. On the other hand, for a higher value of n, we observe a significant decrease in profitability at the second equilibrium, which creates the risk of not resolving the inefficient level of investment.

5.2.2 Equilibrium in a decentralized capacity market with a penalty mechanism

We have now introduced the penalty in the design of the decentralized capacity market. This penalty is meant to punish retailers for selling too many quantities with respect to the level of investment they help provide via the capacity market. While having little interest in the initial case with a price cap (and imperfect competition) generating solely missing money for producers, the penalty takes on its full meaning when we assume inefficient rationing. Indeed, in this case, the additional cost sustained by consumers is due to the inadequacy between the quantity consumed (and hence sold) to consumers and the level of investment. Therefore, the penalty acts as a mean to make partly responsible retailers for this loss.

In terms of market design, the previous observation translates into the fact that the penalty is sustained only when the price cap is binding, that is, only when inefficient rationing occurs. First, let denote an additional threshold value $t_d^w(k)$ such that this is the first state of the world when the price cap is binding after accounting for the depressing effect of the penalty $t_d^w(k) = \{t : p_n(k,t) - S = p^w\}$.[47] We also denote $q_d^w(t)$ the corresponding quantity, which can also be interpreted as the *Cournot* equilibrium in the retail given retailers' profit function and penalty mechanism. With this new value, we can distinguish three cases depending on the value of the installed capacity.

- (Case 1) When $k > q_{0,n}^w(k)$, the price cap is never binding, and the effect of a capacity market on the expected retailers' profit is strictly identical to a regime without a capacity market.
- (Case 2) For a value of k between $q_d^w(k)$ and $q_{0,n}^w(k)$, we observe a paradoxical outcome. Rationing should have occurred as soon as k is below $q_{0,n}^w(k)$ without a penalty. It implies

[47] The penalty does impact retailers at the margin as the mechanism is based on the difference between the realized quantity, retailers' strategic variable, and the investment level.

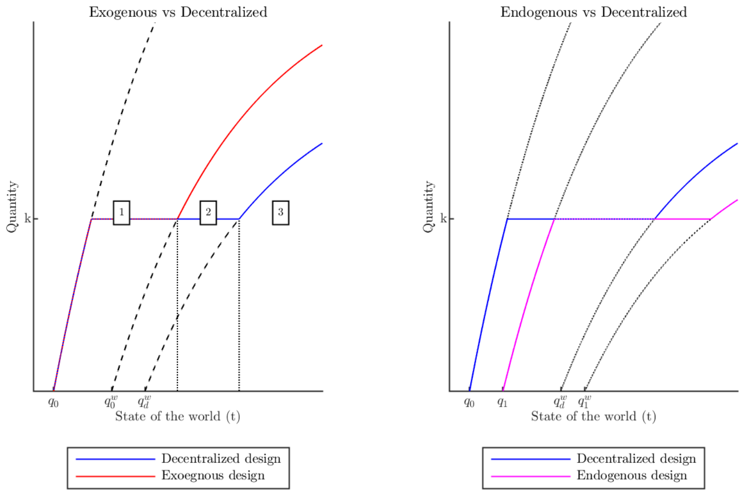


Figure 14: Quantity exchanged on the retail market before rationing

that retailers sustain the penalty, which is then passed to consumers as a marginal cost, which lowers their demand. However, rationing does not occur, which contradicts the demand's decrease due to the penalty. Therefore, retailers follow the level of investment. To do so, they increase the price of their consumers by a unitary amount of $T(k) \leq S$ so that at any state of the world between $t^w_{0,n}(k)$ and $t^w_d(k)$, the demand is equal to the capacity k,[48] that is we have $p^s(q) - T(k) = p^w$.

- (Case 3) k is below $q^w_d(k)$, it is now optimal for the retailers to keep their strategy at $q^w_d(k)$ before rationing as it is the profit-maximizing quantity given the penalty mechanism.

We illustrate those three cases in the following figures. We show how the quantity sold by retailers to final consumers depends on the state of the world before rationing, given the price in the wholesale market and for a level of investment k. The left figure compares the case with an exogenous requirement design (or with no capacity market) and the decentralized capacity market. In Case (1), the decentralized capacity market does not impact retailers. Therefore, the quantity exchanged during off-peak periods at the marginal cost $(q_{0,n}(t))$ and the investment level during on-peak periods (k) when the price cap is not binding is the same. When the price cap starts

[48]We could also assume the reverse mechanism where retailers pay consumers $T(k)$ to reduce the demand in order to avoid the penalty.

binding, retailers should sustain the penalty. However, the mismatch between the incentives in case (2) leads retailers to sell the same quantity as the investment level. On the other hand, without this penalty, retailers should be selling the quantity at the price cap ($q^w_{0,n}(t)$). When this quantity is too important with respect to the penalty, retailers stop lowering their consumption and start offering at the level that maximizes their profit given the penalty mechanism ($q^w_d(t)$). However, due to the negative effect of the penalty, this quantity is still lower than the initial one ($q^w_{0,n}(t)$). The right figure compares the decentralized design with the endogenous case. It illustrates the difference in terms of each design's indirect effect on the system. During off-peak periods, the need for the endogenous market design to allocate the capacity price onto the demand implies a depreciation effect of the quantity exchanged at the price cap ($q_{1,n}(t)$ instead of $q_{0,n}(t)$). In turn, this increases the duration of off-peak periods. Such an indirect effect does not exist in the decentralized capacity market, as the penalty mechanism is applied only when the price cap starts binding. When both designs are in the on-peak periods, and the price cap is not binding, the effect of the price allocation is the same as prices are formed by the demand function. Finally, we illustrate the case when the penalty value is not equal to the capacity price in the endogenous regime. The depressing effect on the quantity exchanged at the price cap ($q^w_d(t)$ and $q^w_{0,n}(t)$) differs. In the figure, model parameters lead to a higher capacity price compared to the penalty. Which implies that $q^w_{1,n}(t) > q^w_d(t)$.

Given the three different cases, the expected profit function of a retailer i becomes:

$$
\begin{aligned}
\Pi^s_d(k) = \quad & \int_0^{t_{0,n}(k)} q_{0,n}(k) m p(q(t)) dF(t) + \int_{t_{0,n}(k)}^{t^w_{0,n}(k)} + k m p(k) dF(t) && \text{Case (1)} \\
& + \int_{t^w_{0,n}(k)}^{t^w_d(k)} k(p(k,t) - p^w - T(k,t)) dF(t) && \text{Case (2)} \\
& + \int_{t^w_d(k)}^{+\infty} k(p(k,t) - p^w) dF(t) - \int_{t^w_d(k)}^{+\infty} S(q^w_d - k) dF(t) && \text{Case (3)} \\
& - p^c(k) k_i
\end{aligned}
$$

It comprises three main parts related to different values of k given a demand level (or a different level of demand given a value of k). The two first terms are the same with and without a capacity market, as the price cap is not binding. The retail price rises while the wholesale price is fixed and equal to the price cap for the second term. Between the two states of the world, $t^w_{0,n}(k)$ and $t^w_d(k)$, the demand decreases due to the retailers' actions to avoid paying the penalty. It is materialized by the transfer $T(k,t)$[49]. When it is not profitable to reduce the demand given the penalty (when $t^w_d(k)$ is reached), then the new demand is given by $q^w_d(k)$, the retailer profit is in the fourth term, and the retailers pay the penalty in the fifth term. The last term is the capacity cost due to the retailer's obligation to buy their capacities. Given this expected profit, we can define the marginal value of a capacity for the retailer, which serves as the retailer's willingness to pay for an additional capacity. Under the market design, the retailers aggregated demand function in the capacity market is equal to the marginal value of an additional capacity for their profit function. Hence, we provide the derivative of retailers' expected profit with respect to k in the following equation.

$$
\begin{aligned}
\frac{\partial \Pi^s_d(k)}{\partial k} = & \int_{t_{0,n}(k)}^{t^w_{0,n}(k)} mp(k) - kmp_q(k))dF(t) + \int_{t^w_{0,n}(k)}^{\infty} (p(k,t) - pw - kp(k))dF(t) \qquad \left(\frac{\partial \Pi^w_d(k)}{\partial k}\right) \\
& + \int_{t^w_{0,n}(k)}^{t^w_d(k)} \left([mp(k) + kmp_q(k)] - [p(k,t) + p_q(k) - p^w]\right) dF(t) \\
& - kSf(t^w_d)\frac{\partial t^w_d(k)}{\partial k} \qquad (\mu(s)) \\
& + \int_{t^w_d(k)}^{\infty} SdF(t)
\end{aligned}
$$

We note $\mu(S)$ the marginal effect induced by the penalty mechanism, including the three last lines. With a penalty, we find first a similar effect of k on the expected profit, which is captured by the first term. The second term represents the effect induced by the penalty during the case (2). When retailers lower their sales by an amount $T(k)$, they essentially get the same expected profit as when

[49] Note that if we assume that retailers pay the consumers to reduce their consumption, only the sign changes

the price cap is not binding, which is represented by a gain both in terms of additional unitary mark-up at the investment level ($mp(k)$) and by a higher mark-up due as it is similar as an increase of demand ($kmp_q(k)$). On the other hand, compared to the initial case without a penalty, they loose the expected marginal profit when the price cap was binding ($p(k,t) + p_q(k) - p^w$). The last terms both relate to the penalty mechanism. The first one comes from the different structure of the revenue when the price cap is binding or not; the probability of being reached changes with k. When k increases, the states of the world when the price case binds are less likely, which is captured by $-\frac{\partial t_d^w(k)}{\partial k}$. Moreover, recall the definition of $t_d^w(k) = \{t : p_n(k,t) - S = p^w\}$, which is also equal to $t_d^w(k) = \{t : p(k,t) - mp(k) - S - p^w = 0\}$, hence we have the following relation between the expected total revenue between $t_{0,n}^w(k)$ and $t_d^w(k)$: $kmp(k) - k(p(k,t) - p^w) = kS$. This value is then expressed in terms of the probability of reaching the threshold $t_d^w(k)$, hence $f(t_d^w)$. Finally, the last term corresponds to the expected unitary penalty sustained during rationing. A marginal increase of k lowers the occurrence of rationing, directly lowering the penalty mechanism's expected cost.

The equilibrium in the capacity market with a penalty mechanism is described in the following Lemma. It is compared to the previous equilibrium without the penalty, and we assess the relative profitability of each possible equilibrium.

Lemma 9. (i) The penalty mechanism described in assumption 5 induces two possible market equilibriums in terms of investment level:

(a) the first one is the same as with the price cap case and without a penalty: k_1^d,

(b) the second one is the investment level such that $\max(k_{0,n}^w, k_3^d)$. With $k3^d$ given by the following equality $k3^d = \{k : \phi_3^d(k) = r\}$, with

$$\phi_3^d(k) = \left[\int_{t_{0,n}(k)}^{t_{0,n}^w(k)} (p(k,t) - kmp_q(k) - c)dF(t) + \int_{t_{0,n}^w(k)}^{\infty} (p(k,t) + kp(k) - c)dF(t)\right] + \mu(S) = r$$

(ii) Moreover, (i) if $\mu(S) > 0$ (resp. $\mu(S) < 0$) then $k_4^d \leq k_2^d$ (resp. $k_4^d \geq k_2^d$). (iii) k_0^d is now relatively more profitable than k_4^d compared to k_2^d.

The first equilibrium has the same interpretation as in the previous lemmas. When the capacity level is sufficiently high, the price cap never binds, and the incentives induced by the penalty mechanism are null. The marginal value of an additional capacity is the same as in the case of a capacity market without a penalty mechanism and without a price cap in the wholesale market. The second equilibrium encompasses the penalty effect compared to the value k_2^d. To see that the first and second terms of the equality in the brackets are equal, the left-hand side of the equation defines k_2^d: $\phi_2^d(k)$. When solving for the equilibrium, we find that the rest of the equality condition is equal to the marginal effect of the penalty mechanism on the expected total profit.[50] Therefore, at the equilibrium, the difference between the two investment levels is given by the net sign of the three last parts. If the net effect of the penalty mechanism is positive, then we have $k_1^d \leq k_2^d$. Otherwise, if it is negative, then $k_2^d \leq k_1^d$.

The first implication of Lemma 9 is that the penalty mechanism has an ambiguous effect on one of the equilibriums. As previously explained, during the case (2), retailers' behavior implies a change in the marginal profit received on the retail market by switching expected profits of on-peak periods when the price cap is binding to expected profits when the price cap is not binding. If the expected profit in the first case is significantly higher than in the second case, it is, therefore, more profitable for retailers to reduce the level of investment. In other words, the marginal cost induced by the penalty of not having enough investment is overshadowed by the loss of expected profit when the price cap is binding. On the other hand, despite not analyzing the coordination issue related to the existence of the two equilibriums, we find that the penalty mechanism always reduces the profitability of the lower equilibrium k_4^d. In the case the first equilibrium k_0^d is more efficient, it can significantly improve the benefit of a capacity market.

[50] This is followed by construction, as the demand function in the capacity market is equal to the derivative of the total expected profit.

5.2.3 Social welfare under uncertain decentralized capacity market

We turn now to the study of the implications of a decentralized capacity market from a social welfare aspect, which needs to be compared to the first-best investment level given by the previous market design regimes, namely the exogenous/endogenous regime and the market share allocation. For clarity, we only analyze the case when we take into account inefficient rationing, as the decentralized capacity market with a penalty mechanism is usually explicitly implemented to account for this inefficiency.

The efficiency of a decentralized relies on two channels: (i) the market equilibrium it provides to the system and (ii) the indirect effect it has on the expected social welfare. The first channel has been partly addressed in the previous analysis, where we characterized the market equilibrium. The second channel starts by defining the new expected welfare function under a decentralized market design and with inefficient rationing.

$$W_d^{bo}(k) = \int_0^{t_{0,n}(k)} \int_0^{q_{0,n}(t)} (p(q,t) - c)dqdF(t) + \int_{t_{0,n}(k)}^{\infty} \int_0^k (p(k,t) - c)dqdF(t) - rk - \int_{t_d^w(k)}^{+\infty} J(\Delta_d k)dF(t)$$

The expression can be rewritten as :

$$W_d^{bo}(k) = W_{0,n}(k) - rk - \int_{t_d^w(k)}^{+\infty} J(\Delta_d k)dF(t)$$

Recall that $W_{0,n}(k)$ is the expected social welfare under the exogenous market design (or without capacity) market. With perfect competition, this value is equal to $W_0(k)$. Therefore, the incentives provided by a decentralized capacity market can be fully, and is only, captured through a change in the cost of inefficient rationing[51]. To see this, we use the model specification provided in example **??**, and we define the cost as follows:

$$M_d(k) = \int_{t_d^w(k)}^{+\infty} \frac{q_d^w(t) - k}{k} \int_0^k (p(q,t) - p^w) dq dF(t)$$

The main changes with the previous designs are found in the new threshold values: $t_d^w(k)$ and $q_d^w(t)$. As previously explained, the incentives created by the penalty mechanism induce retailers to lower their sales in any state of the world between $t_{0,n}^w(k)$ and $t_d^w(k)$ (case(2)), such that the quantity sold is strictly equal to installed capacity. Therefore, no inefficient rationing was implemented during those periods. We summarize in the following proposition the policy implications of a decentralized capacity market with inefficient rationing and in terms of the first-best investment level. We note k_d^* this first-best value such that $k_d^* = \{k : \phi^d(k) = r\}$, with

$$\phi^d(k) = \phi_{0,n} - \int_{t_d^w(k)}^{+\infty} \frac{\partial J(\Delta k)}{\partial k} |_{\Delta k = \Delta_d k} + \frac{J(\Delta_d k)}{\partial \Delta_d k} \frac{\partial \Delta_d k}{\partial k}$$

Proposition 10. (i) A decentralized capacity market always provides higher welfare than under an exogenous market design ($W_d^{bo}(k) \geq W_{0,n}^{bo}(k)$), and the new optimal investment is always lower: $k_d^* \leq k_{0,n}^{bo}$. Moreover, k_d^* and $W_d^{bo}(k)$ are converging respectively towards $k_{0,n}^*$ and $W_0(k)$ as the penalty value S increases.

(ii) (a) A decentralized capacity market always provides higher welfare than under an endogenous market design ($W_d^{bo}(k) \geq W_{1,n}(k)$) if the following inequality holds.

$$\Delta W_{1,n}^{bo}(k) \geq \Delta M_d(k)$$

With $\Delta W_{1,n}^{bo}(k)$ and $\Delta M_d(k) = M_{0,1}(k) - M_d(k)$

[51] A practical implementation issue arising with a decentralized capacity market is how to consider the revenue from the penalty mechanism. On an aggregate level, this should not matter as long as its allocation does not marginally alter the expected surplus. This can be done by assuming a fourth agent in the expected welfare, the government, or by assuming that the penalty revenue is allocated on a lump sum basis to consumers, for instance.

(b) It is sufficient for $M_{1,n}(k) \geq M_d(k)$ to hold so that $\Delta W^{bo}_{1,n}(k) \geq \Delta M_d(k)$ always holds. Moreover, an increase in the penalty value S increases the efficiency of the decentralized capacity market first-best compared to the endogenous market design.

The first statement (i) implies that the decentralized capacity market and for any investment level k (including the first-best investment level) always has a positive benefit in terms of welfare compared to the exogenous market design. Moreover, the increase of the penalty value also always has a positive effect on the expected social welfare, as the only indirect effect it causes is to reduce the inefficient rationing cost. That is, we always have $\frac{\partial M_d(k)}{\partial k} \leq 0$. For sufficiently high penalties, the incentive is such that rationing cost never occurs at the first-best level, hence mimicking a system without inefficient rationing. The second statement (ii) has a natural interpretation: if the gains in terms of lower costs due to the penalty incentives are higher than the net gain of the endogenous regime (comprised of the demand depreciating effect, the decentralized design does not have, and the lower rationing cost), then the decentralized market also provides more expected welfare. Finally, when the rationing cost is lower in the decentralized case than the exogenous one, the efficiency of the former is higher than the efficiency of the latter. Moreover, suppose an increase in the penalty lowers the cost of inefficient rationing. In that case, it reduces the value of $\Delta M_d(k)$ (if it is negative) and can even change its sign so that the decentralized capacity market is always more efficient. In other terms, the second efficiency channel of the decentralized capacity market implies a strictly increasing relation between the expected social welfare and the penalty value. Following those results, we turn now to the relation between the market equilibrium given by the decentralized capacity market.

From an efficiency perspective, it is sufficient to note that a decentralized capacity market is more efficient if the expected social welfare is higher under one of its market equilibrium, not only at its first-best investment.[52] That is, we have either, for instance, $W^{bo}_d(k^d_1) \geq W^{bo}_{1,n}(k^*_{1,n})$ or $W^{bo}_d(k^d_2) \geq W^{bo}_{1,n}(k^*_{1,n})$ so that the decentralized regime is more efficient than a endogenous regime. Unfortunately, the general framework prevents from having a clear-cut answer to when

[52]The coordination between the two possible equilibriums is therefore crucial but left to future works, as one equilibrium can make the decentralized capacity market better or worse than another market design

the decentralized capacity is more efficient[53]. The conditions defined in Proposition 10 for the first best investment level still give the main economic intuitions that can be applied to the relative efficiency of the market equilibrium: if at the decentralized market equilibrium, the reduction in the rationing cost is higher than the net gain of the endogenous market design at the first best investment level then the decentralized capacity market is more efficient. Meanwhile, at the end of this paper, we provide some comparative statistics to analyze the relation between the efficiency of a capacity market and the penalty mechanism that acts as a regulatory tool to manage this efficiency.

Lemma 10. (i) The market equilibrium k_1^d is independent of the penalty, and the market equilibrium k_2^d is a concave function with respect to the penalty.

(ii) The efficiency of a decentralized capacity market : (a) at the market equilibrium k_1^d is strictly increasing with the penalty value. (b) at the market equilibrium k_2^d is strictly increasing with the penalty value if the following inequality holds under the model specification.

$$1 - F(t_d^w(k_2^d)) \geq 2Sf(t_d^w)\frac{\partial t_d^w(k_2^d)}{\partial S}$$

This equilibrium emerges because the price cap never binds at this level, meaning there is no expected penalty cost. However, the strict increasing relation between the expected social welfare under the decentralized capacity market regime and the penalty value implies that an increase in the penalty always increases the expected social welfare at the market equilibrium. On the other hand, the equilibrium k_2^d is concave with respect to the penalty value. This has strong policy implications. For a given set of initial parameters, a maximum constrained expected social welfare can be achieved through the decentralized capacity market. To say it differently, lets denote S^d the penalty threshold such that $S^d = \{S : 1 - F(t_d^w(k_2^d)) = 2Sf(t_d^w)\frac{\partial t_d^w(k_2^d)}{\partial S}\}$. For any value of the penalty $S \in [0, S^d]$, the expected social welfare at the second market equilibrium increases with S, similar to the first equilibrium. However, for any value of the penalty beyond S^d, the expected social welfare is also increasing if and only if the net effect of the decentralized capacity market

[53]Even with the model specification of examples 1 and 4.1.

is positive. This net effect is comprised of (i) a positive effect due to the increase of the expected social welfare implied by the reduction of the rationing cost, as described in the Proposition 10; and (ii) a negative effect due to the reduction of the quantity chosen at the second market equilibrium.

but this maximum does not necessarily ensure that it implies a more efficient decentralized capacity market compared to other market designs.

6 Conclusion and discussion

This paper built a tractable framework to analyze multiple markets' interdependences for an essential good prone to underinvestment, such as electricity or medical supplies. We showed how the investment decisions are affected by those markets, their structure (such as the degree of competition), and, most importantly, their design. Our case study is the capacity markets that were implemented to encourage producers to invest by providing additional remuneration. Most of the literature on capacity markets has focused on the supply side, where producers offer their availability on future transaction periods on the wholesale market. Therefore, the demand side has been overlooked, even though some system efficiency effects are well known. Current implementations show many options regarding the demand side's design on capacity markets, as consumers do not have proper incentives to buy capacities. Using our framework, we compare multiple market designs and their implications. The first set of regimes is based on differentiating the capacity cost allocation. The second set of regimes is represented by how the design can account for current demand realization. We underline the different parameters that can significantly affect the outcomes of a capacity market on investment decisions. The choice of the design can significantly affect prices and quantities in the three markets and the redistribution of welfare between agents. One of the advantages of this framework relies on the possible extensions that we can implement besides providing a simple but complete vision. The rest of the section discusses two issues that could be addressed in future research using this framework.

First, we initially assumed that consumers were fully reactive to retail prices. Such assumptions do not describe the reality yet, as illustrated in the electricity system, as most small final consumers, such as households, are still under fixed-price contracts. The study of final consumers' heterogeneity and its implications for investment decisions in the power system is an emerging trend. Léautier (2014) and Léautier (2016) provide a relevant model close to the one presented in this paper. They show the effects of having those two types of consumers with different investment decisions and a capacity market. However, the author does not compare demand design options for capacity markets and does not consider retailers. Therefore, implementing this new extension in our model could shed light on the issue associated with power systems' investment decisions. It could also significantly impact retailers' individual market design options. Indeed, let's consider that some consumers cannot react to price, but retailers are still forced to cover their consumption. The demand function's formation in the capacity market will be significantly impacted.

Finally, we assume that future consumer demand is commonly shared between agents. A single entity, potentially regulated, and retailers could access a different quantity and quality of information. For instance, we can assume that the entity only has a global vision of future demand, and hence, it is prone to make a more significant error forecast than retailers. On the other hand, retailers have private access to more precise information on their client portfolios while sharing common information on the world's future global states. Therefore, introducing these private/common elements in our model could shed new light on the effect of capacity markets and their market design options. Finally, in some current implementations, the entity based its global forecast on retailers' information. Consequently, the comparison between the various regimes' cases could be analyzed using game theory and signaling.

References

Acemoglu, D., Bimpikis, K., and Ozdaglar, A. (2009). Price and capacity competition. *Games and Economic Behavior*, 66(1):1–26.

Ahuja, A., Athey, S., Baker, A., Budish, E., Castillo, J. C., Glennerster, R., Kominers, S. D., Kremer, M., Lee, J., Prendergast, C., et al. (2021). Preparing for a pandemic: Accelerating vaccine availability. In *AEA Papers and Proceedings*, volume 111, pages 331–35.

Boiteux, M. (1949). *La tarification des demandes en pointe: Application de la théorie de la vente au coût marginal*. place Henri-Bergson.

Brown, D. P. (2018). Capacity payment mechanisms and investment incentives in restructured electricity markets. *Energy Economics*, 74:131–142.

Bublitz, A., Keles, D., Zimmermann, F., Fraunholz, C., and Fichtner, W. (2019). A survey on electricity market design: Insights from theory and real-world implementations of capacity remuneration mechanisms. *Energy Economics*, 80:1059–1078.

Bushnell, J., Flagg, M., and Mansur, E. (2017). Capacity Markets at a Crossroads. *Energy Institute at Haas.* .

Creti, A. and Fabra, N. (2007). Supply security and short-run capacity markets for electricity. *Energy Economics*, 29(2):259–276.

Crew, M. A. and Kleindorfer, P. R. (1976). Peak load pricing with a diverse technology. *The Bell Journal of Economics*, pages 207–231.

De Frutos, M.-A. and Fabra, N. (2011). Endogenous capacities and price competition: The role of demand uncertainty. *International Journal of Industrial Organization*, 29(4):399–411.

De Maere d'Aertrycke, G., Ehrenmann, A., and Smeers, Y. (2017). Investment with incomplete markets for risk: The need for long-term contracts. *Energy Policy*, 105:571–583.

Doorman, G., Barquin, J., Barroso, L., Batlle, C., Cruickshank, A., Dervieux, C., Flanagan, R., Gilmore, J., Greenhalg, J., Höschle, H., et al. (2016). Capacity mechanisms: needs, solutions and state of affairs. *CIGRÉ, Paris.*

Fabra, N. (2018). A primer on capacity mechanisms. *Energy Economics*, 75:323–335.

Fabra, N., Motta, M., and Peitz, M. (2021). Learning from electricity markets: How to design a resilience strategy.

Hobbs, B. F., Hu, M. C., Iñón, J. G., Stoft, S. E., and Bhavaraju, M. P. (2007). A dynamic analysis of a demand curve-based capacity market proposal: The PJM reliability pricing model. *IEEE Transactions on Power Systems*, 22(1):3–14.

Holmberg, P. and Ritz, R. A. (2020). Optimal capacity mechanisms for competitive electricity markets. *The Energy Journal*, 41(Special Issue).

Joskow, P. and Tirole, J. (2007). Reliability and competitive electricity markets. *The Rand Journal of Economics*, 38(1):60–84.

Keppler, J. H., Quemin, S., and Saguan, M. (2021). Why the sustainable provision of low-carbon electricity needs hybrid markets: The conceptual basics.

Klemperer, P. D. and Meyer, M. A. (1989). Supply function equilibria in oligopoly under uncertainty. *Econometrica: Journal of the Econometric Society*, pages 1243–1277.

Léautier, T.-O. (2014). Is mandating. *The Energy Journal*, 35(4).

Léautier, T.-O. (2016). The visible hand: ensuring optimal investment in electric power generation. *The Energy Journal*, 37(2).

Leautier, T.-O. (2018). On the long-term impact price caps: Investment, uncertainty, imperfect competition, and rationing. *International Journal of Industrial Organization*, 61:53–95.

Llobet, G. and Padilla, J. (2018). Conventional power plants in liberalized electricity markets with renewable entry. *The Energy Journal*, 39(3).

Newbery, D. (2016). Missing money and missing markets: Reliability, capacity auctions and interconnectors. *Energy Policy*, 94:401–410.

Ockenfels, A. (2021). Marktdesign für eine resiliente impfstoff-produktion. *Perspektiven der Wirtschaftspolitik*, 22(3):259–269.

Scouflaire, C. (2019). *Capacity remuneration mechanisms: analytical assessment of contemporary experiences and lessons for the future design of electricity markets*. PhD thesis, PSL - Paris Dauphine University.

Vives, X. (1999). *Oligopoly pricing: old ideas and new tools.* MIT press.

Weitzman, M. L. (1974). Prices vs. quantities. *The review of economic studies*, 41(4):477–491.

Wolak, F. A. (2021). Long-term resource adequacy in wholesale electricity markets with significant intermittent renewables. Working Paper 29033, National Bureau of Economic Research.

Yan, H. (2020). Auctions with quantity externalities and endogenous supply. *International Journal of Industrial Organization*, 71:102638.

Zöttl, G. (2011). On optimal scarcity prices. *International Journal of Industrial Organization*, 29(5):589–605.