

# The Visible Hand: Ensuring Optimal Investment in Electric Power Generation

Thomas-Olivier Léautier\*

---

## ABSTRACT

This article formally analyzes the various corrective mechanisms that have been proposed and implemented to alleviate underinvestment in electric power generation. It yields three main analytical findings. First, physical capacity certificates markets implemented in the United States restore optimal investment if and only if they are supplemented with a “no short sale” condition, i.e., producers can not sell more certificates than they have installed capacity. Then, they raise producers’ profits beyond the imperfect competition level. Second, financial reliability options, proposed in many markets, are effective at curbing market power, although they fail to fully restore investment incentives. If “no short sale” conditions are added, both physical capacity certificates and financial reliability options are equivalent. Finally, a single market for energy and operating reserves subject to a price cap is isomorphic to a simple energy market. Standard peak-load pricing analysis applies: under-investment occurs, unless production is perfectly competitive and the cap is never binding.

**Keywords:** Imperfect competition, Market design, Investment incentives

<http://dx.doi.org/10.5547/01956574.37.2.tlea>

## 1. INTRODUCTION

An essential objective of the restructuring of the electric power industry in the 1990s was to “push to the market” decisions and risks associated with investment in power generation, i.e., to have market forces, not bureaucrats, determine how much investment is required, and to have investors, not rate-payers, bear the risks of excess capacity, construction cost overruns and delays.

However, since the early 2000s, generation adequacy has become an issue of concern for policy makers, power System Operators (*SOs*), and economists. It would appear that, contrary to the initial belief, the market does not necessarily provide for the adequate level of generation capacity. Britain, that pioneered the restructuring of the electricity industry in 1990, constitutes the most recent and striking example: Ofgem, the energy regulator recently warned that, in its high demand scenario, involuntary curtailment of customers will be imposed for 8 hours on average during the winter 2015/2016 (Ofgem, 2013).

Operating and regulatory practices aimed at preventing the exercise of market power are often considered to be the primary cause of this market failure. As shown in Marcel Boiteux (1949)’s seminal analysis, high prices a few hours per year are required to finance the optimal capacity. However, in most jurisdictions *SOs* impose *de jure* or *de facto* price caps, that deprive producers

\* Toulouse School of Economics (IAE, IDEI, CRM), 21 allée de Brienne, F31 000 Toulouse. E-mail: thomas.leautier@tse-fr.eu. Tel: + 33 6 33 34 60 68.

of these high prices. This revenue loss, called “missing money”, is considered an important driver of underinvestment in generation (Joskow, 2007).

Therefore, *SOs* and policy makers worldwide have designed and implemented a variety of mechanisms to correct this apparent market failure (Finon and Pignon, 2008). For example, most US power markets have adopted highly structured and prescriptive physical certificates markets, and many European countries are considering, designing or implementing capacity mechanisms.<sup>1</sup>

These mechanisms are extremely complex, hence expensive to set up and run. Furthermore, they constitute a partial reversion towards central planning, which restructuring precisely attempted to eliminate: using a centralized system reliability model, the *SO* sets a generation capacity target, and organizes its procurement. Risk of overcapacity is borne by consumers, while risk of cost overrun is borne by investors. A rigorous economic analysis of the performance of the various market designs implemented by *SOs* to restore investment incentives is therefore required. This is the objective of this article. I am not aware of any previous systematic analytical comparison of these designs.

This work draws on a rich literature, that can be structured along two themes. A first group of articles examines generation investment in restructured power markets. While these works differ in important aspects, most model two stage games: in stage 1, producers decide on installed capacity; in stage 2 they produce and sell in the spot markets, subject to the installed capacity constraint. For example, Borenstein and Holland (2005) and Joskow and Tirole (2007), building on Boiteux (1949) and Crew and Kleindorfer (1976), have developed the benchmark model of optimal investment and production when (i) demand is uncertain at the time the investment decision is made, and (ii) a fraction of the demand does not react to price. The former article considers the perfect competition case, while the latter introduces some elements of imperfect competition. Murphy and Smeers (2005) have developed models of closed- and open-loop Cournot competition at the investment and spot market stages, and characterized the equilibria of these games. Boom (2009) and Boom and Buehler (2014) have examined the impact of vertical integration on equilibrium investment, while Fabra et al. (2011) have examined the impact of the structure of the auction in the spot market on the equilibrium investment. A more recent literature (e.g., Garcia and Shen, 2010) examine multiperiod investment decisions. This article builds on the two-stage Cournot game formalized in Zöttl (2011).

A second group of works describes and analyzes the possible corrective mechanisms.<sup>2</sup> Stoft (2002) discusses average Value of Lost Load (*VoLL*) pricing, Hogan (2005) proposes an energy cum operating reserves markets, and Cramton and Stoft (2006 and 2008) and Cramton and Ockenfels (2011) propose a financial reliability options mechanism.<sup>3</sup> Joskow and Tirole (2007) show that a capacity market and a price cap do not restore the first best with more than two states of the world. Chao and Wilson (2005) examine the impact of options on spot market equilibrium, investment, and welfare. Zöttl (2011) determines the welfare maximizing price cap in the spot market. However, none of these works presents a rigorous comparison of these mechanisms in a general and common setting.

1. France formally instituted a capacity obligation mechanism in March 2012, to be effective in 2015. Britain, Germany, and Belgium are designing mechanisms to ensure adequate capacity.

2. Since these mechanisms are described extensively in the article, they are not developed further here.

3. Strictly speaking, these options ensure “resource adequacy”, not “reliability”. Nevertheless, I use the word “reliability options” as it was the term used in the original Cramton and Stoft articles.

This article bridges these two strands of literature, that analyzes the proposals described in the second group of articles using a rigorous economic model developed in the first group: an extension of the two-stage Cournot model developed by Zöttl (2011) to include both “price reactive” customers and “constant price customers”, the latter being unable to react to spot energy prices and being rationed in some instances (Borenstein and Holland, 2005, Joskow and Tirole, 2007, Stoft, 2002, and Hogan, 2005). Its contribution is to propose clear policy recommendations, building on the economic analysis of these mechanisms. While this work’s primary focus is the electric power industry, the analysis presented here can serve as a basis to examine (under)investment issues in other industries where participants must select capacity in the presence of significant demand variability and uncertainty and limited storage possibilities, for example telecommunications and transport networks.

This article yields three main analytical findings. First, it examines the equilibrium of markets where energy and forward physical installed capacity certificates are separately exchanged. This is the case for example in the Northeast of the United States: 3 to 5 years ahead, the *SO* procures from producers physical capacity certificates (usually 15 to 20% higher than anticipated peak load to protect against supply and demand fluctuations). The cost of these purchases is then passed on to customers. Proposition 1 shows that the *SO* must impose a “no short sale” requirement, i.e., require producers to sell less certificates than have installed capacity (or to build as much capacity as they have sold certificates). If she does, a physical capacity certificates market restores investment incentives: the resulting capacity installed is optimal. For a given price cap, social welfare is thus maximized. However, producers profits are higher than the imperfect competition outcome without the capacity market. Numerical illustration suggests the additional rent from the capacity market is not negligible, that ranges between 10 to 16% of the investment cost.

Second, this article analyzes the equilibrium of another form of forward markets, where producers are required to sell, through the *SO*, financial call options to customers, covering all the demand up to a certain level at a given strike price. Option sellers pay customers the difference between the actual spot energy price and the strike price (Oren, 2005, Cramton and Stoft, 2006 and 2008, Cramton and Ockenfels, 2011).

Proposition 2 proves that options sale reduces but does not eliminate market power. Installed capacity is higher with options sale than without, but still lower than socially optimal. To ensure optimal investment, the *SO* must again impose a “no short sale” requirement. If she does, Proposition 3 shows that financial reliability options and physical capacity certificates with the “no short sale” conditions are equivalent if the “technical” parameters are identical (e.g., if the option strike price equals the wholesale price cap). Reliability options thus also sur-remunerate strategic underinvestment. While Propositions 2 and 3 are consistent with Chao and Wilson (2005) and Allaz and Villa (1993)’s theoretical analysis of the interaction between forward and spot markets, they are new to the literature.

Finally, this articles examines the “energy cum operating reserves market” proposed by Hogan (2005). *SOs* procure operating reserves to protect against an unplanned generation outage. Hogan (2005) proposes the *SO* balances supply against demand for energy *and* operating reserves, using the average *VoLL* as a price cap. Producers should receive additional revenues since: (i) the resulting power price is higher than when the *SO* balances supply against demand for energy alone, and (ii) capacity providing operating reserves—but no energy—is remunerated. This additional revenue is expected to resolve the missing money problem, hence restores investment incentives. However, Proposition 4 shows this intuition is invalid: since these additional revenues are already accounted for in the determination of the installed capacity, the situation is isomorphic to standard peak-load pricing.

Each of these three mechanisms is examined individually in this article, while they may be implemented jointly in practice. For example, most *US* markets have a physical certificate mechanism and co-procurement of energy and operating reserves.

The analysis yields clear policy recommendations. If policy makers and the *SO* are confident a market is sufficiently competitive, as may be the case in Texas, there is no need to impose a price cap and set up a forward capacity market (physical or financial), which are complex and costly to administer. Average *VoLL* pricing or an energy cum operating reserves market are simple to set up and, if the *VoLL* used is close enough to the real *VoLL*, cause limited distortion compared to the optimum. Furthermore, an energy cum operating reserves market remunerates flexibility, an important issue which is not covered in this work.

On the other-hand, policy makers may determine that generation is insufficiently competitive in their jurisdiction. This may be the case in European markets, where in most markets less than 10 generation companies actually compete. This may also be the case where congestion on the transmission grid separates the market in smaller submarkets, and producers may be able to exert local market power. Then, policy makers should set up a forward capacity market as an interim measure while removing barriers to competition. If they believe they can effectively enforce no short sale conditions, they can choose between physical capacity certificates and financial reliability options. If they believe that no short sale conditions are too costly to enforce, they should prefer financial reliability options.

The article is structured as follows. Section 2 presents the model structure and examines the causes of underinvestment. Section 3 examines markets for physical installed capacity certificates. Section 4 analyzes financial reliability options. Section 5 analyzes the “energy cum operating reserves market”. Finally, Section 6 suggests future research directions. Technical proofs are included in the online Appendix.

## 2. UNDERINVESTMENT

The model used throughout this article is developed in Léautier (2014), building on the analysis presented by Zöttl (2011). This Section presents its main features and conclusions. The interested reader is referred to Léautier (2014) for a comprehensive presentation of the model.

### 2.1 Model Structure

#### Uncertainty

Uncertainty is an essential feature of power markets. In this work, demand uncertainty is explicitly modeled, while production uncertainty is taken into account implicitly through operating reserves (presented in Section 5). This representation is suitable for markets that rely mostly on controllable generation technologies, such as thermal and nuclear (see for example Chao and Wilson, 1987). Extension to markets where intermittent sources constitute an important portion of the generation portfolio is discussed in the concluding section.

The number of possible states of the world is infinite, and these are indexed by  $t \in [0, +\infty)$ . The functions  $f(t)$  and  $F(t)$  are respectively the *ex ante* probability and cumulative density functions of state  $t$ . Since all market participants have the same information about future demand projections and construction plans,  $f(t)$  and  $F(t)$  are common to all stakeholders.

## Supply

This article considers a single generation technology, characterized by marginal cost  $c > 0$  and investment cost<sup>4</sup>  $r$ . A single technology is sufficient to analyze total installed capacity, that depends solely on the characteristics of the marginal technology (see for example Boiteux (1949) for the perfect competition case and Zöttl (2011) for the imperfect competition case).

## Underlying demand

**Assumption 1.** *All customers have the same underlying demand  $D(p,t)$  in state  $t$ , where  $p$  the electric power price, up to a scaling factor.*

Assumption 1 greatly simplifies the derivations, while preserving the main economics insights. Inverse demand is  $P(q,t)$  defined by  $D(P(q,t),t) = q$ , and gross consumers surplus is  $S(p,t) = \int_0^{D(p,t)} P(q,t) dq$ .  $P(q,t)$  is downward sloping:  $P_q(q,t) < 0$ . States of the world are ordered by increasing demand:  $P_t(q,t) > 0$ .

## Constant price customers, curtailment, and Value of Lost Load

Only a fraction  $\alpha > 0$  of customers face and react to real time wholesale price (“price reactive” customers), while the remaining fraction  $(1 - \alpha)$  of customers face constant price  $p^R$  in all states of the world (“constant price” customers).

Since a fraction of customers does not react to real time price, there may be instances when the *SO* has no alternative but to curtail demand, i.e., to interrupt supply. As discussed for example in Joskow and Tirole (2007), there exists multiple rationing technologies. Curtailment is represented by a *servicing ratio*  $\gamma \in [0,1]$ :  $\gamma = 0$  represents no servicing (i.e., all energy to all consumers is curtailed), while  $\gamma = 1$  represents full servicing (i.e., no customer is curtailed).  $\mathcal{D}(p,\gamma,t)$  is the demand for price  $p$  and servicing ratio  $\gamma$  in state  $t$ ,  $\mathcal{P}(q,\gamma,t)$  is the inverse demand for a given servicing ratio  $\gamma$ , defined by  $\mathcal{D}(\mathcal{P}(q,\gamma,t),\gamma,t) = q$ , and  $\mathcal{S}(p,\gamma,t)$  is the gross consumer surplus.

**Assumption 2.** *The SO has the technical ability to curtail “constant price” consumers while not curtailing “price reactive” consumers.*

Assumption 2 holds only partially today: most *SOs* can only organize curtailment by geographical zones, and cannot differentiate by type of customer. However, most price reactive customers are large enough that they are connected directly to individual transformers or to the high voltage grid, hence they need not be curtailed when the *SO* curtail constant price customers. Assumption 2 will hold fully in a few years, when smart grids are rolled out, as is mandated in most European countries and many US states. It is expected that *SOs* will then be able to differentiate among adjacent customers on the basis of the information provided by power suppliers, and curtail constant price customers individually.

4. Both are expressed in €/MWh.  $r$  is the annual capital cost expressed in €/MWh/year divided by 8,760 hours. It includes the cost of risk.

Since customers are homogeneous, the *SO* has no basis for discriminating among them. Rationing is thus proportional, i.e., the fraction  $(1 - \gamma)$  of constant price customers is completely cut-off for the duration of the outage.

When consumers are curtailed, the marginal Value of Lost Load (*VoLL*) represents the value they place on an extra unit of electricity (Joskow and Tirole, 2007, Stoft, 2002), formally defined as

$$v(p, \gamma, t) = \frac{\frac{\partial \mathcal{S}}{\partial \gamma}}{\frac{\partial \mathcal{D}}{\partial \gamma}}(p, \gamma, t).$$

If the *SO* knew the *VoLL* for every rationing technology and state of the world (and each customer class), the second best (as defined in the next Section) would be achieved.<sup>5</sup> In reality, regulators, *SOs* and economists have little idea of the *VoLL*. Estimation is extremely difficult, because the *VoLL* varies drastically across customer classes, states of the world, and duration and conditions of outages. Estimates vary in an extremely wide range from 2,000 £/*MWh* in the British Pool in the 1990s to 200,000 \$/*MWh* (see for example Cramton and Lien, 2000, and Praktiknjo and Erdmann, 2012). In practice, the *SO* uses her best estimate of the average *VoLL*, i.e., the average value per *MWh* of electricity lost for an average customer who loses all his service, and prioritizes curtailment by geographic zones (economic activity, political weight, network conditions, etc.), thus implementing a third best.

Both approaches produce downward sloping demand curves, hence are analytically equivalent. In this work, I assume the *SO* knows exactly the *VoLL*. While this assumption is unrealistic, it constitutes a useful analytical benchmark.

## 2.2 Socially Optimal Consumption and Investment

### Optimal consumption

The residual inverse demand curve with possible curtailment of constant price customers is

$$\rho(Q, t) = P\left(\frac{Q - (1 - \alpha)\mathcal{D}(p^R, \gamma^*, t)}{\alpha}, t\right), \quad (1)$$

where  $\gamma^*$  is the optimal serving ratio in state  $t$  for production  $Q$ .

Price reactive customers face the wholesale spot price  $\rho(Q, t)$ , hence are never curtailed at the optimum. Off-peak, demand is low, and production  $Q(t)$  is determined by  $\rho(Q(t), t) = c$ . On-peak, demand is set by installed capacity  $K$ , and the wholesale price is  $\rho(K, t)$ .

As long as  $\rho(K, t) \leq v(p^R, 1, t)$ , constant price customers are not curtailed in state  $t$ . If  $\rho(K, t) > v(p^R, 1, t)$ , then  $\gamma^* < 1$  is set to equalize constant price customers' *VoLL* and the wholesale price

5. The *VoLL* would be different if instead all price responsive customers received only  $(1 - \gamma)$  of their demand. However, this does not change the analytical treatment presented in this article. I am grateful to a referee for this observation.

$$v(p^R, \gamma^*, t) = \rho(K, t).$$

Define  $\bar{t}(K)$  the first state of the world when curtailment may occur.<sup>6</sup> If curtailment never occurs,  $\bar{t}(K) \rightarrow +\infty$ . With a slight abuse of notation, define  $\rho_q = \frac{1}{\alpha} P_q \left( \frac{Q - (1 - \alpha)D(p^R, t)}{\alpha}, t \right)$  if no rationing occurs, and  $\rho_q = \frac{\partial v}{\partial K} = \frac{\partial v}{\partial \gamma} \frac{\partial \gamma^*}{\partial K}$  if rationing occurs. Léautier (2014) derives sufficient conditions for  $\rho(Q, t)$  to be well-behaved, even when curtailment occurs.

As an illustration, suppose (i) inverse demand is linear with constant slope:  $P(q, t) = a(t) - bq$ , and (ii) rationing perfectly anticipated and proportional:  $S(p, \gamma, t) = \gamma S(p, t)$  and  $D(p, \gamma, t) = \gamma D(p, t)$ . If no rationing occurs,

$$\rho(Q, t) = \frac{a(t) - bQ - (1 - \alpha)p^R}{\alpha}.$$

Since rationing is anticipated and proportional,

$$v(p^R, \gamma, t) = \frac{S(p^R, t)}{D(p^R, t)} = a(t) - b \frac{D(p^R, t)}{2} = \frac{a(t) + p^R}{2}.$$

### Optimal investment

The marginal social value capacity is

$$\Psi(K, c) = \int_{\hat{t}_0(K, c)}^{+\infty} (\rho(K, t) - c) f(t) dt,$$

where  $\hat{t}_0(K, c)$  is the first state of the world such that price (weakly) exceeds the marginal cost for production  $K$

$$\rho(K, \hat{t}_0(K, c)) \geq c.$$

$\Psi(K, c)$  is decreasing in both arguments. If  $\rho(0, 0) > c + r$ , the optimal capacity  $K^*$  is the unique solution to

$$\Psi(K^*, c) = r.$$

Off-peak, as long as capacity is not constrained, price equals marginal cost, hence marginal capacity generates no economic profit. On-peak, when capacity is constrained, price exceeds marginal cost. The optimal capacity is set such that the marginal social value capacity is exactly equal to the marginal capacity cost  $r$ .

If  $\alpha$  is small, rationing of constant price customers may occur at the optimal capacity, an

6.  $\bar{t}$  is a function of all the parameters. The notation  $\bar{t}(K)$  is used since the dependency on installed capacity  $K$  is the most important in this analysis.

issue known as the Theoretical (capacity) Adequacy Problem (*TAP*). With the specification summarized in Appendix A, Léautier (2014) finds that rationing occurs at the optimal capacity until  $\alpha = 3.9\%$  if the price elasticity of demand  $\eta = -0.01$ , and  $\alpha = 13.9\%$  if  $\eta = -0.1$ . This result may seem counter-intuitive: a less elastic demand results in less curtailment! The intuition is that, for a given  $\alpha$ , capacity is higher when demand is more inelastic, hence, curtailment is less frequent.

If the presence of constant price customers was the only imperfection in power markets, an energy only market design, sometimes referred to as average *VoLL* pricing, would be efficient (Stoft, 2002, Oren, 2005): when constant price customers are curtailed, the *SO* pays energy at the *VoLL*. This yields optimal investment, conditional on the *VoLL*. If the *SO* knew exactly the *VoLL*, this would achieve a second best. Otherwise, this would yield a third best.

However, power markets are subject to other imperfections. First, competition among producers is less than perfect. Second, producers may be risk averse, which reduces their investment. Finally, investment decisions are dynamic and long-lived, more complex than a simple static model suggests. This article focusses on the first imperfection, that examines the performance of corrective mechanisms in a static model where agents are risk neutral. Extensions to a dynamic model and risk-averse agents are left for further work.

### 2.3 Imperfect Competition, Price Cap, and Underinvestment

Consider now  $N$  producers, that play a two-stage game: in stage 1, producer  $n$  installs capacity  $k^n$ ; in stage 2 he produces  $q^n(t) \leq k^n$  in the spot market in state  $t$ . Producers are assumed to compete à la Cournot in the spot markets, facing inverse demand  $\rho(Q, t)$  defined by equation (1). Stage 2 can be interpreted as a repetition of multiple states of the world over a given period (for example one year), drawn from the distribution  $F(\cdot)$ .

Producers are assumed to be independent, i.e., not to be vertically integrated into supply. A rich literature has examined how vertical integration modifies market behavior, in particular investment incentives (Boom, 2009, and Boom and Buehler, 2014).

The game is solved by backwards induction: producers first compute profits from a Nash equilibrium in the energy spot market for each state of the world  $t$ , given installed capacities  $(k^1, \dots, k^N)$ ; then they make their investment choice in stage 1 based on the expectation of these spot market profits.

Aggregate production in state  $t$  and aggregate installed capacity are respectively  $Q(t) = \sum_{n=1}^N q^n(t)$  and  $K = \sum_{n=1}^N k^n$ . Producer's  $n$  profit for the two-stage game is  $\Pi^n(k^n, \mathbf{k}^{-n})$ .

The results presented in this article hold for other forms of imperfect competition in the spot market, as long they yield an equilibrium price higher than the marginal cost  $c$ , and a profit function  $\Pi^n(k^n, \mathbf{k}^{-n})$  with the required concavity. Cournot competition is used as it provides simple analytical expressions that can be illustrated numerically.

To limit the exercise of market power, the *SO* imposes a cap  $\bar{p}^W$  on the wholesale power price,<sup>7</sup> assumed to satisfy

$$c + r \leq \bar{p}^W \leq \rho(0, 0).$$

7. In practice, most SOs in the United States impose a cap on *bids* into the wholesale markets, not a cap on wholesale price. A wholesale price cap simplifies the analysis, while preserving the main economic insights.



A price cap lower than the full marginal cost of the first unit of energy would block any investment. A cap higher than the value of the first unit of energy consumed would have limited effectiveness.

The first on-peak state of the world under imperfect competition, i.e., where the marginal revenue for production  $K$  equals marginal cost is  $\hat{t}(K,c,N)$ , uniquely defined by

$$\rho(K, \hat{t}(K,c,N)) + \frac{K}{N} \rho_q(K, \hat{t}(K,c,N)) = c.$$

The aggregate capacity constraint may be binding before or after the price cap constraint in the relevant range, i.e.,  $\hat{t}(K,c,N) \leq \hat{t}_0(K, \bar{p}^W)$  or  $\hat{t}(K,c,N) > \hat{t}_0(K, \bar{p}^W)$ . Introducing constant price customers makes  $\hat{t}(K,c,N) > \hat{t}_0(K, \bar{p}^W)$  a distinct possibility, in particular if the residual demand  $\rho(Q,t)$  is very inelastic, i.e., if  $\alpha$  or  $|\eta|$  are very low.

Léautier (2014) proves that, if certain technical sufficient conditions are met, the equilibrium capacity  $K^C(\bar{p}^W)$  is characterized by

$$\Omega(K^C, \bar{p}^W) = r,$$

where  $\Omega(K, \bar{p}^W)$  is defined by

$$\Omega(K, \bar{p}^W) = \left( \int_{\hat{t}(K,c,N)}^{\hat{t}_0(K, \bar{p}^W)} \left( \rho(K,t) + \frac{K}{N} \rho_q(K,t) - c \right) f(t) dt \right) I_{\{\hat{t}(K,c,N) \leq \hat{t}_0(K, \bar{p}^W)\}} + \int_{\hat{t}_0(K, \bar{p}^W)}^{+\infty} (\bar{p}^W - c) f(t) dt$$

where  $I_{\{x \geq 0\}}$  is the indicator function, that takes the value 1 if  $x \geq 0$  and 0 otherwise.

This result illustrates the two distortions that reduce investment. First, if generation produces at capacity before the cap is reached, imperfect competition reduces the marginal value of capacity by two terms: the reduction in profit on the inframarginal units as in all Cournot competition models  $\left( \frac{K}{N} \int_{\hat{t}(K,c,N)}^{\hat{t}_0(K, \bar{p}^W)} \rho_q(K,t) f(t) dt \right)$ , but also the lost margin  $(\rho(K,t) - c)$  in the states of the world  $t \in [\hat{t}_0(K,c), \hat{t}(K,c,N)]$ . Both effects are negative. Second, whether the cap or the generation capacity constraint is reached first, the price cap reduces the marginal value, since the  $SO$  values energy at  $\rho(K,t)$ , while producers receive only  $\bar{p}^W < \rho(K,t)$ . This is the “missing money” discussed for example by Joskow (2007), and Cramton and Stoft (2006).

Léautier (2014) then computes the resulting capacity, and proposes sufficient conditions for the existence of price cap that maximizes welfare. The latter result extends Zöttl (2011) result to the presence of constant price customers ( $\alpha \in (0,1)$ ). Taking these constant price customers into account yields welfare maximizing price caps that are much higher than those observed in most markets. Thus existing price caps will lead to underinvestment, hence the need for corrective mechanisms.

### 3. PHYSICAL CAPACITY CERTIFICATES

The  $SO$  imposes price cap  $\bar{p}^W$  on the energy markets and procures at least  $K^*$  physical capacity certificates from producers. To simplify the notation and analysis, operating reserves are ignored: as will be proven in Section 5, including them would not modify the economic insights. All units (old and new) receive the same compensation in the physical certificates markets.

The timing is as follows:

1. The *SO* designs the rules of the energy and capacity markets. All parameters are set
2. Producers sell physical capacity certificates to the *SO*, according to the rules set up previously
3. Producers build new capacity if needed
4. The spot markets are played. In each state, producers compete à la Cournot facing  $\rho(Q,t)$ , given their installed capacity and their physical capacity obligation. The *SO* pays the physical certificates to the producers, and passes the cost of purchase to customers.

To simplify the analysis, this pass-through is assumed not to distort consumption decisions in the spot market, e.g., the pass-through is proportional to the size of the meter.<sup>8</sup>

Steps 2 and 3 can be inverted or simultaneous: generators first build the plants, then sell physical capacity certificates, or build and sell simultaneously.<sup>9</sup>

$\phi^n$  and  $\Phi = \sum_{m=1}^N \phi^m$  are respectively the certificates sold by producer  $n$  and the aggregate volume of certificates sold. In practice, *SOs* offer a “smoothed” (inverse) demand curve:

$$H(\Phi) = \begin{cases} r & \text{if } \Phi \leq K^* \\ h(\Phi) & \text{if } K^* < \Phi < K^* + \Delta\bar{K} \\ 0 & \text{if } \Phi \geq K^* + \Delta\bar{K} \end{cases}$$

where (i)  $r$ , the capital cost of capacity, is the maximum price the *SO* is offering for capacity, (ii)  $\Delta\bar{K} > 0$  is an arbitrary capacity increment, and (iii)  $h(\cdot)$  is such that  $H(\cdot)$  is  $C^2$ , except maybe at  $K^*$  and  $K^* + \Delta\bar{K}$ ,  $h'(\Phi) < 0$ ,  $2h'(\Phi) + \phi h''(\Phi) < 0$  for all  $\phi$ , and

$$|h'(K^*)| \geq \frac{Nr}{K^*}. \quad (2)$$

As will be discussed below, condition (2) simplifies the exposition, but is not essential. It is met in practice. For example, Cramton and Ockenfels (2011) suggest a linear form for  $h(\cdot)$  with  $\frac{\Delta\bar{K}}{K^*} = 4\%$ . Condition (2) is then equivalent to  $N \frac{\Delta\bar{K}}{K^*} \leq 1$ , and holds as long as less than 25 producers compete.

Efficiency of the physical certificates market is conditional on the quality of the *SO's* estimate of the optimal capacity  $K^*$ . The three stage game (stages 2, 3, and 4) is solved below, assuming the *SOs* knows perfectly  $K^*$ , uses  $r$  as the reserve price, and the function  $H(\cdot)$  is given. Determining the optimal stage–1 parameters is left for further work. The equilibrium characterized as follows:

8. One least distorting approach would be a levy proportional to peak consumption (in  $kW$ ). Since these payments are assumed not to distort consumption decisions, they have no impact on social welfare. This requires smart meters to measure peak consumption. I am grateful to a referee for this observation.

9. The formal proof can be found in a previous version of this article, available at <http://idei.fr/doc/wp/2012/visible.pdf> <http://idei.fr/doc/wp/2012/visible.pdf>

**Proposition 1.** *The SO must impose and monitor that the installed capacity exceeds the capacity certificates sold by each generator:  $k^n \geq \phi^n$ . Then (i) producers issue as many credits as they install capacity, and (ii)  $K^*$  is the unique symmetric equilibrium investment level. Compared to the no installed capacity market situation, producer's profit and overall welfare are increased.*

*Proof.* The full proof is presented in Appendix B. Existence of a physical capacity certificates market alone does not alter investment incentives. The SO must impose  $k^n \geq \phi^n$ , otherwise  $K^C$  remains the installed capacity.

If she does, producers sell exactly as many certificates as they have installed capacity since incremental capacity is unprofitable unless it collects capacity markets revenues. Then, since  $k^n = \phi^n$  at the equilibrium, producer  $n$  program is:

$$\max_{k^n} \Pi_{CM}^n(k^n, \mathbf{k}_{-n}) = \Pi^n(k^n, \mathbf{k}_{-n}) + k^n H(K)$$

Given the shape of the inverse demand function  $H(\cdot)$ ,  $k^n = \frac{K^*}{N}$  for all  $n$  is the unique symmetric equilibrium, and producers' profit is:

$$\Pi_{CM}^n\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right) = \Pi^n\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right) + \frac{K^*}{N} r.$$

Then, since  $\Pi^n(k, \dots, k)$  is concave (Zöttl, 2011) and  $K^C \leq K^*$ ,

$$\Pi^n\left(\frac{K^C}{N}, \dots, \frac{K^C}{N}\right) \leq \Pi^n\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right) + \left(\frac{K^C - K^*}{N}\right) \frac{\partial \Pi^n}{\partial k}\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right)$$

$\Leftrightarrow$

$$\begin{aligned} \Delta &= \Pi_{CM}^n\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right) - \Pi^n\left(\frac{K^C}{N}, \dots, \frac{K^C}{N}\right) \geq -\frac{K^C}{N} \frac{\partial \Pi^n}{\partial k}\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right) \\ &\quad + \frac{K^*}{N} \left(\frac{\partial \Pi^n}{\partial k}\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right) + r\right) > 0 \end{aligned}$$

since  $\frac{\partial \Pi^n}{\partial k}\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right) < 0$  and  $\frac{\partial \Pi^n}{\partial k}\left(\frac{K^*}{N}, \dots, \frac{K^*}{N}\right) + r = \Omega(K^*) > 0$ .

Producers' profits increase compare to the no installed capacity market situation. Finally, since overall welfare  $W(K)$  increases up to to  $K = K^*$ ,  $W(K^*) \geq W(K^C)$ . ■

Capacity markets do not automatically restore investment incentives. In the model, producers exercise market power by reducing capacity *ex ante*, and not by withholding output on-peak. The SO must therefore ensure that producers cannot sell short, i.e., sell more certificates than their installed capacity.

This observation is not original to this work, for example it has been articulated by Wolak (2006). Yet it remains an important practical challenge for SOs, that monitor that existing generation assets providing certificates are still operational, and that planned capacity having received certificates has indeed be installed. SOs then impose a penalty on producers that, when requested, do

not offer in the spot market energy up to the certificates they have sold forward. The implementation challenge lies in assessing when commitment has not been met, hence this process is still evolving. For example, *ISO* New England recently proposed new rules for its forward market to ensure producers have incentives to produce. The ban on short-selling is not universal: demand-side resources can effectively sell-short in most *US* markets.

Physical capacity markets increase overall welfare, and also increase transfers from customers to producers. This result is very general. Denote  $K^E$  (not necessarily equal to  $K^*$ ) the equilibrium capacity including the certificates markets. As long as  $\Pi^n(K, \dots, K)$  is concave, and  $K^E > K^C$ , the marginal value of capacity for the producers at  $K^E$  is negative:  $\frac{\partial \Pi^n}{\partial k} \left( \frac{K^E}{N}, \dots, \frac{K^E}{N} \right) < 0$ . The equilibrium price in the capacity market ( $r$  in this case) must compensate for this negative marginal value, otherwise  $K^E$  would not be an equilibrium:  $\frac{\partial \Pi^n}{\partial k} \left( \frac{K^E}{N}, \dots, \frac{K^E}{N} \right) + r \geq 0$ . This is sufficient for the proof.

Although I had never seen its formal proof, this result is intuitive: producers must receive a rent to induce them to invest beyond the oligopoly capacity. The illustrative model developed by Léautier (2014) provides an estimate of this additional rent  $\Delta$ . It varies slightly with the price cap  $\bar{p}^W$  and the proportion of price reactive customers  $\alpha$ . To simplify, I provide the average value of  $\Delta$  over all relevant values of  $\bar{p}^W$  and for  $\alpha = 5\%$ , which appears appropriate for most markets. For price elasticity of demand  $\eta = -0.01$  the average rent is around 5,100 €/MWh, approximately 10% of investment cost; for  $\eta = -0.1$  the average rent is around 8,400 €/MWh, approximately 16% of investment cost. These estimates illustrate that the rent created by the capacity market is not trivial.

Is there an optimal structure to the physical certificates market? With the above design, the only parameter that can be modified in the price cap  $\bar{p}^W$ . I choose as an objective function the net surplus from consumption, therefore transfers from consumers to producers do not impact social welfare. Then, since the resulting capacity is  $K^*$  for all  $\bar{p}^W$ , the latter has no impact on the resulting capacity. However, increasing  $\bar{p}^W$  always increases welfare, as it reduces the probability of curtailment: when the cap is binding, no customer can respond to price, hence the *SO* must curtail. Thus, there exists no optimal binding price cap with a capacity market as modeled here.

Finally, if condition (2) is not met, the aggregate capacity at the unique symmetric equilibrium is  $K_{CM}^C \in (K^*, K^* + \Delta \bar{K}]$ . Welfare increases if and only if  $\Delta \bar{K}$  is small enough that  $W(K^* + \Delta \bar{K}) \geq W(K^C)$ .

#### 4. FINANCIAL RELIABILITY OPTIONS

Financial contracts constitute another mechanism used in power markets. This Section examines financial reliability options, proposed by Oren (2005), Cramton and Stoft (2006 and 2008), and more recently Cramton and Ockenfels (2011). Options and not forward contracts are the financial instruments analyzed here, since Chao and Wilson (2005), that examine a slightly different option design, argue that options are in general preferable. These options constitute an insurance against spot energy prices higher than a pre-agreed strike price  $\bar{p}^S$ , sold by producers to customers. If the spot price  $p(t)$  is lower than  $\bar{p}^S$ , producer  $n$  does not make any payment. If  $p(t) > \bar{p}^S$ , producer  $n$  pays  $(p(t) - \bar{p}^S)$  times a fraction of the realized demand equal to his fraction of the total options sale.

The *SO* does not impose a cap on wholesale prices, and runs an auction for financial reliability options.  $\theta^n$  and  $\Theta = \sum_{m=1}^N \theta^m$  are respectively the options sold by producer  $n$  and the

aggregate volume of options sold. To limit the potential exercise of market power, Cramton and Ockenfels (2011) propose the *SO* impose all capacity must be committed forward through option sales:  $\theta^n \geq k^n$ .

The timing and notation are identical to the capacity market case, except that the subscript  $_{RO}$  is added when appropriate. A very simple auction setup is assumed, similar to the one suggested by Cramton and Stoft (2008): the *SO* determines the volume she desires to purchase, assumed to be  $K^*$ , sets the capital cost of capacity  $r$  as the reserve price for the auction, and proposes a downward sloping inverse demand curve for options:

$$H_{RO}(\Theta) = \begin{cases} r & \text{if } \Theta \leq K^* \\ h_{RO}(\Theta) & \text{if } K^* < \Theta < K^* + \Delta\bar{K}_{RO} \\ 0 & \text{if } \Theta \geq K^* + \Delta\bar{K}_{RO} \end{cases}$$

where (i)  $\Delta\bar{K}_{RO} > 0$  is an arbitrary capacity increment, and (ii)  $h_{RO}(\cdot)$  is such that  $H_{RO}(\cdot)$  is  $C^2$ , except maybe at  $K^*$  and  $K^* + \Delta\bar{K}$ ,  $h'_{RO}(\phi) < 0$ ,  $2h'_{RO}(\phi) + \phi h''_{RO}(\phi) < 0$  for all  $\phi$ , and  $h_{RO}(\cdot)$  verifies condition (2).

When the spot price exceeds the strike price, price-reactive consumers then pay  $\bar{p}^S$  as the effective price, i.e., they know when making their consumption decision they receive rebate  $\max(\rho(Q, t) - \bar{p}^S, 0)$  per unit of energy purchased. Then, actual demand does not depend on the spot price, which leads to rationing.

$\hat{t}_0(K, \bar{p}^S)$  is the first state of the world such that the spot price exceeds the strike price, and is defined by  $\rho(K, \hat{t}_0(K, \bar{p}^S)) \geq \bar{p}^S$ . We assume  $\bar{p}^S$  satisfies

$$\Psi(K^C(\bar{p}^S), \bar{p}^S) \leq r. \tag{3}$$

Condition (3) simplifies the exposition, as it guarantees that  $\Theta = K^*$  is the unique equilibrium of the options market, however it is not essential. As shown in Appendix C,  $\Psi(K^C(p), p)$  is decreasing in  $p$ , and  $\Psi(K^C(p), p) \rightarrow 0$  as  $p$  stops binding. Thus, condition (3) is met for  $\bar{p}^S$  sufficiently high.

Chao and Wilson (2005) examine a slightly different market structure: they consider physical options paired (or not) with a complementary price insurance, and compute the linear supply function equilibrium for options forward sales and power spot sales. Their findings are aligned with those presented below.

#### 4.1 Expected Profits with Financial Reliability Options

The producers profit function is characterized below:

**Lemma 1.** *The expected profit of producer  $n$  is*

$$\Pi^n_{RO}(k^n, \theta^n, \mathbf{k}_{-n}, \theta_{-n}) = \theta^n H_{RO}(\Theta) + \Pi^n(k^n, \mathbf{k}_{-n}) + \left(k^n - \frac{\theta^n}{\Theta} K\right) \Psi(K, \bar{p}^S), \tag{4}$$

with the convention that  $\bar{p}^S$  acts as the price cap in  $\Pi^n$ .

*Proof.* Producer  $n$  receives the revenues from options sale  $\theta^n H_{RO}(\Theta)$ , plus profits from the energy market. Suppose first  $\hat{t}(K, c, N) \leq \hat{t}_0(K, \bar{p}^S)$ . First, the producer receives profit  $\Pi^n(k^n, \mathbf{k}_{-n})$  previously

computed, assuming  $\bar{p}^S$  as price cap. Second, since there is no price cap, he receives the difference between the spot price  $\rho(K,t)$  and the cap  $\bar{p}^S$  for every unit produced when the price exceeds  $\bar{p}^S$ . Since  $\hat{i}(K,c,N) \leq \hat{i}_0(K,\bar{p}^S)$ , he produces his entire capacity  $k^n$ , hence he receives  $k^n \int_{\hat{i}_0(K,\bar{p}^S)}^{+\infty} (\rho(K,t) - \bar{p}^S) f(t) dt = k^n \Psi(K,\bar{p}^S)$ .

Finally, when the spot price exceeds the strike price  $\bar{p}^S$ , each generator must pay  $(\rho(K,t) - \bar{p}^S)$  times his fraction  $\frac{\theta^n}{\Theta}$  of the total demand. Since  $\hat{i}(K,c,N) \leq \hat{i}_0(K,\bar{p}^S)$ , total demand is equal to total capacity  $K$  and the payment is proportional to  $\frac{\theta^n}{\Theta} K$ . Total expected payment from generator  $n$  is thus:  $\frac{\theta^n}{\Theta} K \int_{\hat{i}_0(K,\bar{p}^S)}^{+\infty} (\rho(K,t) - \bar{p}^S) f(t) dt = \frac{\theta^n}{\Theta} \Psi(K,\bar{p}^S)$ . Summing these terms yields equation (4).

Appendix C proves that Equation (4) also obtains if  $\hat{i}_0(K,\bar{p}^S) < \hat{i}(K,c,N)$ . ■

The profit realized in states higher than  $\hat{i}_0(K,\bar{p}^S)$  is  $\pi_{RO}^n(K,t) = k^n \left( \left( 1 - \frac{\theta^n}{\Theta} \frac{K}{k^n} \right) \rho(K,t) + \frac{\theta^n}{\Theta} \frac{K}{k^n} \bar{p}^S - c \right)$ . Producers face a weighted average of the spot price and the option price, hence are less sensitive to an increase in spot price. Consistent with Allaz and Villa (1993) and Chao and Wilson (2005), a producer holding forward contracts faces lower incentives to exert market power in the spot market.

#### 4.2 Equilibrium Capacity with Financial Reliability Options

**Proposition 2.** *Suppose the SO imposes  $\theta^n \geq k^n$  and chooses strike price  $\bar{p}^S$  such that condition (3) holds. Reliability options reduce but do not eliminate the underinvestment problem.  $K_{RO}^C$ , the unique symmetric equilibrium of the options and investment game, verifies*

$$K^C(\bar{p}^S) \leq K_{RO}^C < K^* ,$$

with equality occurring when  $N = 1$ .

*Proof.* Appendix C proves that, if producers invest first then sell options, there exists a unique symmetric equilibrium of the two-stage game, that satisfies

$$\theta^n = \frac{K^*}{N} ,$$

and  $k^n = \frac{K_{RO}^C}{N}$ , where  $K_{RO}^C$  is uniquely defined by

$$\Omega(K_{RO}^C, \bar{p}^S) + \frac{N-1}{N} \Psi(K_{RO}^C, \bar{p}^S) = r. \tag{5}$$

Then,

$$\Omega(K_{RO}^C, \bar{p}^S) = r - \frac{N-1}{N} \Psi(K_{RO}^C, \bar{p}^S) \leq r.$$

Hence  $K_{RO}^C \geq K^C(\bar{p}^S)$ . Then,

$$\begin{aligned} \frac{\partial \Pi_{RO}^n}{\partial k^n} \left( \frac{K^*}{N}, \dots, \frac{K^*}{N} \right) &= - \int_{i_0(K,c)}^{i_1(K,c,N)} (\rho(K^*, t) - c) f(t) dt + \frac{K^*}{N} \int_{i_0(K,c,N)}^{i_1(K,\bar{p}^S)} \rho_q(K^*, t) f(t) dt \\ &\quad - \frac{1}{N} \Psi(K^*, \bar{p}^S) \\ &< 0. \end{aligned}$$

Then  $K^* > K_{RO}^C$  since we prove in Appendix C that  $\Pi_{RO}^n(\frac{K}{N}, \dots, \frac{K}{N})$  is concave.<sup>10</sup> ■

For  $N > 1$ , reliability options curb the exercise of market power: the resulting installed capacity is higher than the Cournot capacity. Thus, they are more effective than physical certificates alone, that have no impact on installed capacity without the “no short sale” obligation.

However, reliability options are not sufficient to completely eliminate market power and restore optimal investment incentives. This result may appear surprising, since reliability options impose a penalty of  $(\rho(K, t) - \bar{p}^S)$  on each unit a producer is short energy. However, a closer examination of the mechanism reveals that, at the symmetric equilibrium, this penalty represents only  $\frac{N-1}{N}(\rho(K, t) - \bar{p}^S)$ , which is not sufficient to fully compensate for the “missing money”  $(\rho(K, t) - \bar{p}^S)$ .

Proposition 2 mirrors Allaz and Villa (1993) analysis of the interaction between spot and forward markets: assuming Cournot competition in both, they show that introducing forward markets reduces but does not eliminate market power, and has not impact on a monopoly ( $N = 1$ ).

Finally, observe that generators short-sell certificates in equilibrium:  $\theta^n = \frac{K^*}{N} \geq \frac{K_{RO}^C}{N}$ .

### 4.3 Equivalence Between Physical Certificates and Financial Reliability Options when “No Short Sale” Conditions are Added

If the *SO* cannot impose a no short sale condition, Proposition 2 above proves that financial reliability options yield higher investment. Which one should the *SO* choose if she can impose a no short sale condition? Proposition 3 below shows that both mechanisms are equivalent, if the technical parameters are equivalent:

**Proposition 3.** *Suppose (i) the SO imposes and monitors that the installed capacity equals the options sold by each generator:  $k^n = \theta^n$ , (ii) the wholesale price cap in the capacity market is set*

10. A previous version of this work, available at <http://idei.fr/doc/wp/2012/visible.pdf> shows that the result also holds if producers sell certificates, then invest, or sell certificates and invest simultaneously.

equal to the strike price of the reliability option ( $\bar{p}^S = \bar{p}^W$ ) and satisfies condition (3), and (iii) the demand functions for reliability options and for capacity credits are identical and satisfy condition (2). Then, financial reliability options yield the same equilibrium as a capacity market with a no short-sale condition.

*Proof.* Since the SO imposes  $\theta^n = k^n$ , equation (4) yields

$$\Pi_{RO}^n(k^n, \mathbf{k}_{-n}) = \Pi^n(k^n, \mathbf{k}_{-n}) + k^n H_{RO}(K).$$

If  $\bar{p}^S = \bar{p}^W$  and  $H_{RO}(\cdot) = H(\cdot)$ , then, with the no short sale conditions,  $\Pi_{RO}^n = \Pi_{CM}^n$ . Thus the equilibria are identical. ■

As mentioned earlier, since producers sell exactly as many options as their installed capacity (or install as much capacity as they sold options), the profit net of the payment on the option is equivalent to a cap on prices. Therefore, if the technical parameters are identical, both approaches are equivalent.

## 5. ENERGY CUM OPERATING RESERVES MARKET

SOs must secure operating reserves to protect the system against catastrophic failure. Hogan (2005) suggests that remuneration of these operating reserves may help resolve the missing money problem.<sup>11</sup>

The representation of operating reserves is that of Borenstein and Holland (2005). For simplicity, only one type of reserves is considered, the non-spinning one (i.e., plants that are not running, but can start up and produce energy within a short pre-agreed time frame). Since the plant is not running, the marginal cost of providing reserves is normalized to zero. In reality, SOs run multiple markets for operating reserves, for example, spinning, 10-minutes, 30-minutes. The economic insights are not modified, as long as the no-arbitrage condition presented below holds.

Hogan (2005) proposes that the SO runs a single market for energy and operating reserves. Generating units called to produce receive the wholesale price  $w(t)$ , generating units that provide operating reserves receive the wholesale price  $w(t)$  less the marginal cost of generation  $c$ , assumed to be perfectly known by the SO. Generators are therefore indifferent between producing energy or providing reserves, an essential condition (Borenstein and Holland, 2005). When an unscheduled generation outage occurs, operating reserves produce energy and receive the full price  $w(t)$ .

Operating reserves requirements are expressed as a percentage of demand, denoted  $h(t)$ ,

11. Hogan (2005) conjectures: "Under stressed conditions there would not be adequate capacity to meet all load and maintain the target nominal level of reserves. This would give rise to scarcity pricing determined by the capacities of the generation offered, the energy demand, and the administrative demand for operating reserves. (. . .) Payments for operating reserves would be made to generators providing reserves and the cost would be applied to loads in a proportional uplift payment. All generators providing energy would receive the high energy price. All generators providing reserves would receive this high energy price less the variable cost of the marginal reserve capacity. Although scarcity conditions with very high prices would apply in relatively few hours, the payments to generators during these hours would include a large fraction of the total contribution to fixed and investment costs."

12. In practice, various metrics for operating reserves are used, including absolute values expressed in MW. Expressing reserves as a percentage of peak demand simplifies the analysis while preserving the main economic intuition.



and taken as given here.<sup>12</sup> Defining the optimal  $h(t)$  requires advanced network analysis, hence is beyond the scope of this work. Joskow and Tirole (2007) show the optimal reserve ratio increases with the state of the world; hence  $h(t)$  is assumed to be nondecreasing.

The retail price  $p(t)$  must be higher than wholesale price  $w(t)$  to cover generators' revenues from the operating reserves market. A natural choice is to directly include the cost of reserves in the retail price faced by price reactive customers:<sup>13</sup>

$$p(t) = w(t) + h(t)(w(t) - c)$$

⇔

$$p(t) - c = (1 + h(t))(w(t) - c) \tag{6}$$

Throughout this section, the retail and wholesale prices are assumed to be related by equation (6). The notation and model structure are identical to the previous Sections, except that the subscript or superscript *OR* is added when appropriate.

Only the fraction  $\frac{1}{1 + h(t)}$  of installed capacity is used to meet demand in state  $t$ , hence  $\frac{K}{1 + h(t)}$  and not  $K$  is the output appearing in the function  $\rho(\cdot, t)$  (a formal proof is presented in Appendix D). Thus, the marginal social value of capacity in state  $t$  is

$$w(K, t) - c = \frac{p(t) - c}{1 + h(t)} = \frac{\rho\left(\frac{K}{1 + h(t)}, t\right) - c}{1 + h(t)}.$$

The marginal social value of capacity is

$$\Psi_{OR}(K) = \int_{\hat{t}_0^{OR}(K, c)}^{+\infty} \frac{\rho\left(\frac{K}{1 + h(t)}, t\right) - c}{1 + h(t)} f(t) dt,$$

where  $\hat{t}_0^{OR}(K, c)$  is uniquely defined<sup>14</sup> by  $\rho\left(\frac{K}{1 + h(t)}, \hat{t}_0^{OR}(K, c)\right) = c$ . The socially optimal capacity is thus uniquely defined by

$$\Psi_{OR}(K_{OR}^*) = r. \tag{7}$$

Consider now the producers' problem. By construction, producers are indifferent between producing energy or providing reserves. In state  $t$ , they offer  $s^n(t)$  into the energy cum operating reserves

13. Borenstein and Holland (2005) show it to be the perfect competition outcome.

14. Since  $h(t)$  is nondecreasing,  $m_1(K; t) = \rho\left(\frac{K}{1 + h(t)}, t\right)$  is increasing in  $t$ :  $\frac{\partial m_1}{\partial t} = -\rho_q \frac{Kh'(t)}{(1 + h(t))^2} + \rho_t > 0$ .

market.  $S(t) = \sum_{n=1}^N s^n(t)$  is the total offer. Energy available to meet demand is  $Q(t) = \frac{S(t)}{1+h(t)}$ . The SO then (i) verifies that  $s^n(t) \leq k^n$ , and (ii) allocates each  $s^n(t)$  between energy  $q^n(t)$  and reserves  $b^n(t)$ . Producer  $n$  profit is then

$$\pi^n(t) = (q^n(t) + b^n(t))(w(t) - c) = \frac{sn(t)}{1+h(t)} \left( \rho \left( \frac{S(t)}{1+h(t)} \right) - c \right),$$

since (i) energy and operating reserves receive same net revenue by construction, and (ii) wholesale ( $w(t)$ ) and retail  $\left( \rho \left( \frac{S(t)}{1+h(t)} \right) \right)$  prices are linked by equation (6). The problem is then isomorphic to standard peak load pricing, except that  $\frac{s^n(t)}{1+h(t)}$  replaces production  $q^n(t)$ .  $\hat{t}^{OR}(K, c, N)$ , the first on-peak state of the world under imperfect competition, is uniquely defined<sup>15</sup> by  $\rho \left( \frac{K}{1+h(\hat{t}^{OR})}, \hat{t}^{OR} \right) + \frac{1}{N} \frac{K}{1+h(\hat{t}^{OR})} \rho_q \left( \frac{K}{1+h(\hat{t}^{OR})}, \hat{t}^{OR} \right) = c$ . The SO imposes a wholesale price cap  $\bar{v}$  equal to her best estimate of *VoLL*.  $\hat{t}_0^{OR}(K, \bar{v})$ , the first state of the world where the cap may be binding, is uniquely defined by  $\rho \left( \frac{K}{1+h(\hat{t}_0^{OR})}, \hat{t}_0^{OR} \right) = \bar{v}$ . For simplicity,  $\bar{v}$  is assumed to be binding after the capacity constraint under imperfect competition:  $\hat{t}^{OR}(K, c, N) \leq \hat{t}_0^{OR}(K, \bar{v})$ . The inverse demand function for producers is then:  $\rho \left( \frac{K}{1+h(t)}, t \right)$  as long as price cap is not reached, and a horizontal inverse demand at  $\bar{v}$  afterwards. Following the steps of the standard peak load analysis, the marginal value of capacity for a producer at the symmetric equilibrium is

$$\begin{aligned} \Omega_{OR}(K) &= \int_{\hat{t}_0^{OR}(K, \bar{v})}^{\hat{t}^{OR}(K, c, N)} \left( \rho \left( \frac{K}{1+h(t)}, t \right) + \frac{1}{N} \frac{K}{1+h(t)} \rho_q \left( \frac{K}{1+h(t)}, t \right) - c \right) f(t) dt \\ &\quad + \int_{\hat{t}_0^{OR}(K, \bar{v})}^{+\infty} \frac{\bar{v} - c}{1+h(t)} f(t) dt, \end{aligned}$$

and there exists a unique symmetric equilibrium for which each generator invests  $\frac{K_{OR}^C}{N}$  defined by:

$$\Omega_{OR}(K_{OR}^C) = r. \tag{8}$$

**Proposition 4.** *Suppose the SO runs an energy cum operating reserves market and imposes a price cap  $\bar{v}$ . The problem is isomorphic to standard peak load pricing.  $K_{OR}^C < K_{OR}^*$  unless (i) generation is perfectly competitive ( $N \rightarrow +\infty$ ), and (ii) the price cap is never binding ( $\hat{t}_0^{OR}(K, \bar{v}) \rightarrow +\infty$ ).*

15. Similarly,  $m_2(t) = \rho \left( \frac{K}{1+h(t)}, t \right) + \frac{1}{N} \frac{K}{1+h(t)} \rho_q \left( \frac{K}{1+h(t)}, t \right)$  is increasing in  $t$  since  $m_2'(t) = - \left( \frac{N+1}{N} \rho_q + \frac{1}{N} \frac{K}{1+h(t)} \rho_{qq} \right) \frac{Kh'(t)}{(1+h(t))^2} + \rho_t > 0$ .

*Proof.* The result follows immediately from equations (8) and (7). ■

Including an operating reserve market leads to the same investment incentives as average *VoLL* pricing. This result is surprising: one would have expected the operating reserves market to alleviate the missing money problem, since (i) all producing units receive a higher price, and (ii) units providing capacity but not energy are remunerated.

However, the discussion above shows these two effects are already included in the determination of the socially and privately optimal capacities  $K_{OR}^*$  and  $K_{OR}^C$ . Then, units providing reserve capacity receive the same profit ( $w(t) - c$ ) as units producing electricity, to avoid arbitrage between markets. No additional profit is generated. The operating reserves market remunerates reserves, which are needed, not capacity investment.

## 6. CONCLUSION

This article formally analyzes the various corrective mechanisms that have been proposed and implemented to alleviate underinvestment in electric power generation. It yields three main analytical findings. First, physical capacity certificates markets implemented in the United States restore optimal investment if and only if they are supplemented with a no short sale condition, i.e., producers can not sell more certificates than they have installed capacity. Then, they raise producers' profits beyond the imperfect competition level. Second, financial reliability options, proposed in many markets, are effective at curbing market power, although they fail to fully restore investment incentives. If no short sale conditions are added, both physical capacity certificates and financial reliability options are equivalent. Finally, a single market for energy and operating reserves subject to a price cap is isomorphic to a simple energy market. Standard peak-load pricing analysis applies: under-investment occurs, unless production is perfectly competitive and the cap is never binding.

This analysis highlights the limitations of the corrective mechanisms. This suggests that policy makers should first and foremost control and reduce the exercise of market power, then use these mechanisms as interim remedial measures. If they believe they can effectively enforce no short sale conditions, they can choose between physical capacity certificates and financial reliability options. If they believe that no short sale conditions are too costly to enforce, they should prefer financial reliability options.

These results provide a sound basis for policy makers decision making. Different avenues for further work would increase their applicability. First, expand the economic models to other types of technologies: (i) intermittent and uncontrollable production technologies such as photovoltaic and on- and off-shore wind mills, which will provide an increasingly important share of power supply;<sup>16</sup> (ii) reservoir hydro production, which has almost zero marginal cost, but limited overall production capacity, and (iii) voluntary curtailment, i.e., consumers reducing their consumption upon the *SO*'s request.

Second, determine the optimal design parameters. In this work, we have taken the auction parameters (reserve price, shape of the demand function) as given. Optimally selecting these parameters may lead to higher welfare.<sup>17</sup>

16. As a first step, Green and Léautier (2014) derive an extension of the peak-load pricing model when subsidized renewables are included.

17. I am grateful to a referee for that suggestion.

Third, examine how these designs perform under different vertical industry structures. Bushnell et al. (2008) have examined how vertical industry structure impacts market performance, and Boom (2009) and Boom and Buehler (2014) have examined how vertical industry structure impacts investment. It would be important to determine which combination of market structure and capacity mechanism is most effective.

Finally, expand the model to multiple investment periods. Observation suggests the power industry, like many capital-intensive industries, displays cycle of over- and under-investment (“boom bust” cycles). Understanding how various market designs perform in a dynamic setting is therefore extremely important.

## ACKNOWLEDGMENTS

I am grateful to Claude Crampes, Bill Hogan, Patrick Mouttapa, Marie-Anne Plagnet, David Salant, Jean Tirole, three anonymous reviewers, and participants to the 2010 *IDEI* conference on the Economics of Energy Markets for insightful discussions and comments on previous drafts, and for a grant from Electricité de France that supported this research. All errors and omissions are my responsibility.

## REFERENCES

- Allaz, B. and J.-L. Vila (1993). “Cournot competition, futures markets and efficiency.” *Journal of Economic Theory* 59: 1–16. <http://dx.doi.org/10.1006/jeth.1993.1001>.
- Boiteux, M. (1949). “La tarification des demandes en pointe: Application de la théorie de la vente au cout marginal.” *Revue Générale d’Electricité* 58: 321–40.
- Boom, A. (2009). “Vertically integrated firm’s investments in electricity generating capacities.” *International Journal of Industrial Economics*, 27: 544–551. <http://dx.doi.org/10.1016/j.ijindorg.2008.08.007>.
- Boom, A. and S. Buehler (2014). “Vertical structure and the risk of rent extraction.” Copenhagen Business School Working Paper 11-2006 (previous version).
- Borenstein, S. (2005). “The long-run efficiency of real time pricing.” *The Energy Journal* 26(3): 93–116. <http://dx.doi.org/10.5547/ISSN0195-6574-EJ-Vol26-No3-5>.
- Borenstein, S. and S. Holland (2005). “On the efficiency of competitive electricity markets with time-invariant retail prices.” *RAND Journal of Economics* 36: 469–493.
- Bushnell, J., E. Mansur, and C. Saravia (2008). “Vertical arrangements, market structure and competition.” *American Economic Review* 98(1): 237–266. <http://dx.doi.org/10.1257/aer.98.1.237>.
- Chao, H. and R. Wilson (1987). “Priority service: Pricing, investment, and market organization.” *American Economic Review* 77: 899–916.
- Chao, H. and R. Wilson (2005). “Resource adequacy and market power mitigation via option contracts.” Technical Report 1010712, Electric Power Research Institute. <http://stoft.com/metaPage/lib/Chao-Wilson-2003-04-resourceadequacy-options.pdf>.
- Cramton, P. and J. Lien (2000). “Value of lost load.” Mimeo, University of Maryland.
- Cramton, P. and A. Ockenfels (2011). “Economics and design of capacity markets for the power sector.” Mimeo, University of Maryland. <http://www.cramton.umd.edu/papers2010-2014/cramton-ockenfels-economicsand-design-of-capacity-markets.pdf>.
- Cramton, P. and S. Stoft (2006). “The convergence of market designs for adequate generation capacity.” White Paper for the California Electricity Oversight Board.
- Cramton, P. and S. Stoft (2008). “Forward reliability markets: Less risk, less market power, more efficiency.” *Utilities Policy* 16(3): 194–201. <http://dx.doi.org/10.1016/j.jup.2008.01.007>.
- Crew, M. and P. Kleindorfer (1976). “Peak load pricing with a diverse technology.” *The Bell Journal of Economics* 7(1): 207–231. <http://dx.doi.org/10.2307/3003197>.
- Fabra, N., N.-H. V. der Fehr, and M.-A. de Frutos (2011). “Market design and investment incentives.” *The Economic Journal* 121: 1340–1360. <http://dx.doi.org/10.1111/j.1468-0297.2011.02450.x>.

- Finon, D. and V. Pignon (2008). "Capacity mechanisms in imperfect electricity markets." *Utilities Policy* 16(3). <http://dx.doi.org/10.1016/j.jup.2008.01.006>.
- Garcia, A. and Z. Shen (2010). "Equilibrium capacity expansion under stochastic demand growth." *Operations Research* 58(1): 30–42. <http://dx.doi.org/10.1287/opre.1090.0719>.
- Green, R. and T.-O. Léautier (2014). "Do costs fall faster than revenues? renewable electricity subsidies dynamics." Presentation available at <http://idei.fr/doc/conf/eem/conf2014/Slides/Green.pdf>.
- Hogan, W. (2005). "On an "energy only" electricity market design for resource adequacy." Mimeo, Center for Business and Government, John F. Kennedy School of Government, Harvard University.
- IEA (2010). Projected Cost of Generating Electricity. OECD/IEA.
- Joskow, P. and J. Tirole (2007). "Reliability and competitive electricity markets." *RAND Journal of Economics* 38(1):60–84.
- Joskow, P. L. (2007). "Competitive electricity markets and investment in new generating capacity." In: Helm, D. (Ed.), *The New Energy Paradigm*, Oxford University Press. <http://dx.doi.org/10.1111/j.1756-2171.2007.tb00044.x>.
- Léautier, T.-O. (2014). "The 'demand side' effect of price caps: Uncertainty, imperfect competition, and rationing." Mimeo, Toulouse School of Economics.
- Lijesen, M. (2007). "The real-time price elasticity of electricity." *Energy Economics* 29: 249–258. <http://dx.doi.org/10.1016/j.eneco.2006.08.008>.
- Murphy, F. and Y. Smeers (2005). "Generation capacity expansion in imperfectly competitive restructured electricity markets." *Operations Research* 53(4): 646–661. <http://dx.doi.org/10.1287/opre.1050.0211>.
- Ofgem (2013). Electricity capacity assessment report 2013. <https://www.ofgem.gov.uk/ofgem-publications/75232/electricity-capacityassessment-report-2013.pdf>.
- Oren, S. (2005). "Ensuring generation adequacy in competitive electricity markets." In J. Griffin and S. Puller, editors, *Electricity Deregulation: Choices and Challenges*, 388–415. University of Chicago Press.
- Praktiknjo, A. J. and G. Erdmann (2012). "Costs of blackouts and supply security in the context of climate protection and nuclear energies in germany." Mimeo, TU Berlin, Department of Energy Systems.
- Stoft, S. (2002). *Power System Economics*. Wiley-Interscience.
- Wolak, F. (2004). "What's wrong with capacity markets?" <http://stoft.com/metaPage/lib/WolaK-2004-06-contract-adequacy.pdf>. Mimeo, Department of Economics, Stanford University.
- Zöttl, G. (2011). "On optimal scarcity prices." *International Journal of Industrial Organization* 29(5) :589–605. <http://dx.doi.org/10.1016/j.ijindorg.2011.01.002>.



Connect with  
**IAEE**  
on facebook

