Market power in electricity markets, suboptimal investment and ISO procurement

Guy MEUNIER & Dominique FINON
Centre International de Recherche sur l'Environnement et le Développement, CNRS &Ecole des Etudes en Sciences Sociales, Paris[1]

In a context of highly concentrated electricity industries and entry barriers, governments may worry that incumbent firms strategically under-invest in generation. Associated with the well known short-term strategy of production restriction, suboptimal investment allows firms to increase price and profits, and retain long-term market power. We analyse a policy response using an ISO procurement approach to restore the long-term social optimum, whereby ISOs have an obligation to meet minimum invest levels in generation capacity and electricity supply.

A dynamic three-stage game is modelled to analyse the strategic interactions between the benevolent ISO and the firms. We analyse first the situation of a monopoly, then of an oligopoly of firms. Two stages of investment, (the first by the firms, the second by the ISO), and a stage of production allow to distinguish long-term and short-term market power. We analyse how the short-term market power of firms could prevent the ISO from restoring the long-term optimum and how the ISO intervention modifies the long-term firms' incentives to invest. We establish that this system can improve the situation but for high concentration of the industry and inelastic demand, the firms are still able to get strictly positive profits

1. Introduction

Ensuring enough generation capacity to meet future electricity demand progressively became a contentious issue in the design of the reforms, particularly since some recent crisis. One motivation behind the reforms is to encourage efficient investment and to avoid situations of overcapacity observed in the former regime of regulated monopoly. However, now there are growing concerns about the ability of the new liberalised market regime to invest sufficiently in building capacity required to ensure supply security avoid long periods of high prices.

Three main reasons are highlighted in the literature, to explain the lack of investment in generation. The first refers to market power. Because of the inelastic nature of demand, even a slight shortfall of available capacity can provoke a dramatic increase in price. Every generator regardless of their market share can potentially benefit from a lack of investment in generation in the system, which will provoke profitable periods of price spikes. Hence generators spur under-investment, instead of aggressively

[1] Centre International de Recherche pour l'Environnement et le Développement,
32 avenue de la Belle Gabrielle, 94736 Nogent sur Marne, France,
Tel. : 33 1 4394 7384 ; Fax : 33 1 4394 7370 ; e.mail : meunier@centre-cired.fr, finon@centre-cired.fr,

expanding their market shares. This is argued as one of the main explanations concerning the lack of generation capacity during the Californian electricity crisis (Borenstein, 2002).

The second reason is the exceptional volatility of electricity price and the corresponding difficulty of risk management, in particular for the development of futures and forward contracts. Such volatility is explained by the short-run inelasticity of supply and demand, and the non-storabile nature of electricity: “On a market on which consumers cannot react to prices in a situation of severe capacity tension, there are no limits to the prices that the producers can fix when a shortfall appears” (R. Green, 2001). The main problem is that it is not possible to distinguish between scarcity rent and market power effects. Every generator, even the smallest, has the ability to make the prices rise by restricting its production.

This volatility is particularly problematic in two ways. First, infra-marginal rents during the peak periods are needed to ensure the profitability of both peak and base load investments, yet these rents are highly uncertain. Moreover the price spikes which are necessary to attract investment in “peakers” raise problems of public acceptability of reforms. Even if large price spikes are efficient signals of scarcity, the public suspicion of market power lead regulators to impose price caps. This adds to the disincentive effects of volatility to discourage investment in “peakers” technologies. Secondly there is a risk that markets overreact to recurrent price spikes and over-invests. This results from uncoordinated investment decisions and may result in a dramatic price drop and bankruptcy of new entrants, as observed in US regional markets after 2000 with “boom and bust” cycles.

These specific risks added to a high regulatory risk discourage investment in generation by risk-averse agents. Long-term contracts are thus necessary for risk adverse generators to invest in capital intensive facilities. In addition, not only the producers but also the buyers are risk averse (Neuhoff et De Vries, 2004; Finon, 2006; Joskow, 2006). In particular, as retail suppliers are not allowed to sell electricity to final consumers by long-term contracts, they do not want to sign long-term contract with producers. This limits the development of a long-term forward market. As such, producers’ vertical integration into the supply business is the dominant arrangement to allow investment in capital intensive equipment.

The third reason argued in the literature is the existence of a public good. This concerns mainly investment in peakers, which includes sufficient reserve capacity to face every random situation on generation equipments and demand. Securing sufficient peak capacity is essential to guarantee reliable supply in the long-run. However, incentives to invest in peak units do not take into account the social and economic advantage to bring sufficient capacity to the system, while every consumer benefits from private decision to invest in securing this capacity. However, it is not technically possible to have private “insurance contracts” that offer a guarantee of supply in relation to the willing to pay of the different consumers.

To compensate underinvestment, the two possible solutions are either public coordination to develop reserve capacity or market instruments to restore incentives to invest in peakers capacity. Market instruments may take form of capacity payments, capacity obligation combined with capacity markets, or reliability options (Oren, 2003, 2005; Perez-Ariaga et al.; 2001; Stoft & Crampton, 2006; Joskow, 2006). Literature on these instruments has mainly focused on three approaches. The first concerns an empirical comparison of instruments along different criteria: effectiveness to contribute to the public good supply, or preventing gaming problems, etc. (De Vries, 2004). The second approach proposes new incentive instruments (Soft and Crampton, 2006; Oren, 2005; Rivier et al, 2003). The third focuses on the economic efficiency of specific instruments (Creti & Fabra, 2004; Joskow & Tirole, 2004).

In this paper we analyse the efficiency of the instrument of ‘procurement by the Independent System Operator (ISO)’. Under this system the regulator mandates the ISO to develop generation capacity, to prevent sub-optimal investment in the system. This instrument is already used in some European countries (Sweden, France) for the development of peak units and reserve capacities and could be used for the development of base load equipments which contribute to the reduction of negative externalities as renewable energy plants and nuclear technology (Finon, 2006).

We do not address here the problem of public good but focus on the issue of market power. Procurement by a benevolent ISO should compensate the suboptimal investment resulting from the exercise of long-term market power. The two issues - market power and public good aspects - are linked. The exercise of long-term market power could amplify the deficit of investment for supplying the public good.

To our knowledge there are few previous analyses of this type on evaluation of instruments. The most developed is the Joskow & Tirole's exercise (2004) on ISO's procurement. The authors analyse in a competitive environment the distortion due to out-of-merit dispatching and cost recovery through an uplift. In their setting, the competitive assumption ensures that the market could realise the long-term optimum. The ISO's intervention is not justified and its distortion effects are analysed.

We tackle here the issue of the procurement by ISO from another perspective: the imperfect competition which leads producers to under-invest, and the possibility for a benevolent regulator to counter the oligopolistic game by the threat to mandate the ISO to develop generation capacity either by himself or by auctioning for long-term contract at a fixed price. To simplify, we assimilate regulator and ISO under the hypothesis of common benevolence. We make the particular hypothesis that the ISO sells the electricity that it produces rather than producing it "out of dispatching" and recovering its costs by an uplift, because the instrument must not add distortions.

We use the methodology of theoretical two-stage oligopoly models around investment along several contributions including von der Fehr and Harbord (1997), Murphy and Smeers (2002) and Boom (2002, 2003). They study investments in new generation capacities assuming that after investment has been decided, all future decisions and information are subsumed into one future period representing the decision of production and the short-term market. We extend their dynamic structure by modelling the situation as a three-stage game: first we model decisions of investment of producers, then decisions of investment by the ISO and then decisions of production firms and the ISO.

The ISO's intervention improves the situation but for inelastic demand and concentrated industry it does not restore the long-term optimum. The two main points are:

- The short-term strategic behaviour of the firms prevents the ISO from reaching the long-term optimum. If the ISO invests in too much capacity, the firms restrict their production and there is a public loss due to this restriction. This phenomenon limits the ISO's investment.
- Firms anticipate the ISO's intervention. In some cases they can gain strictly positive profit by investing in sufficient capacity to put the ISO's in the dilemma described above.

We begin by analysing the situation of a monopoly and a benevolent ISO. We proceed by backward induction to characterize subgame perfect equilibria. In the second part we generalize the model to a situation of oligopoly.

2. A game between an ISO investor and a monopolist.

We first analyse a three-stage game between a monopolist and a benevolent ISO. We consider only one technology and one state of demand (the situation would be similar if we introduce lower states of demand and corresponding technologies and assume competitiveness in these states). In standard models, the regulatory authority induces incentives for producers or consumers (via taxes, price cap, price sell, quotas, contracts…). Here, the ISO acts as a generator who can directly sell electricity on the market. The cost of producing electricity is divided into two parts: a capacity cost I, and an operating cost c. The demand is represented by the inverse function $p(q)$ with q the electricity consumed. We assume that this function is strictly decreasing and is log concave[1], and that there exist positive quantities k^* and k^{**} such that $p(k^*)=c+I$ and $p(k^{**})=c$.

The three different stages of the game are : first the monopolist invests in k_1 units of capacity, then the ISO invests in k_{ISO} units of capacity, and finally they produce the quantities of electricity q_1 and q_{ISO} respectively subject to the constraints $q_1 \leq k_1$ and $q_{ISO} \leq k_{ISO}$. The profit of the monopolist is :

$$\pi_1 = \left(p(q) - c \right).q_1 - I.k_1$$

As the ISO sells electricity on the market, its profit is:

$$\pi_{ISO} = (p-c)q_{ISO} - Ik_{ISO}$$

The ISO is benevolent and maximises the social surplus $W(q,k)$ subject to equilibrium constraint in the production stage. The social surplus is :

$$W(q,k) = \int_0^q p(u)du - c.q - I.k$$

We characterize perfect equilibriums subgame. So we proceed by backward induction. We begin to solve the production stage given fixed capacities, then we analyse the ISO investment decision and finally the firm investment decisions.

2.1. The production stage.

At this stage capacities are fixed. The firm and the ISO choose production quantities. A Nash equilibrium of the game is an intersection of the two reaction functions. A reaction function specifies the optimal output of a firm for each fixed output of its opponent.

The firm maximises her profit and the ISO maximises social surplus. By considering the firm's production q_1 as fixed, the ISO produces $q_{ISO}(q_1,k_{ISO})$ such that: $p(q_1+q_{ISO}) = c$ if $q_{ISO} \leq k_{ISO}$. So the reaction function of the ISO is:

$$q_{ISO}(q_1,k_{ISO}) = \min\left\{k^{**} - q_1, k_{ISO}\right\}$$

The monopolist maximises his profit subject to capacity constraint :

$$\max_{q_1} p(q_{ISO}+q_1)q_1 - cq_1, \text{ s.c. } q_1 \leq k_1$$

In order to describe the firm's reaction function we introduce the following function $r(.)$

$$\forall q_{ISO} \geq 0, r(q_{ISO}) = \arg\max_{q \geq 0} p(q+q_{ISO})q - c.q \quad (1)$$

This function is the short-term reaction function when the firm has a capacity sufficiently high to be never binding. It is a continue differentiable function on the set $[0, k^{**}]$. It satisfies the first order condition, that marginal revenue and short-term marginal cost c are equalised:

$$p + p'r = c, \forall q_{ISO} \in \left[0, k^{**}\right]$$

Its slope[2] is strictly set between 0 and -1 , and $r(k^{**})=0$, which simply means that the firm does not produce when the price is equal to the marginal production cost c. The reaction function of the firm with fixed capacity can be described using r:

$$\forall q_{ISO} \in \left[0, k^{**}\right], q_1(q_{ISO},k_1) = \min\left\{k_1, r(q_{ISO})\right\} = \begin{cases} k_1 \text{ if } \mathrm{q_{ISO}} \leq r^{-1}(k_1) \\ r(q_{ISO}) \text{ otherwise} \end{cases} \quad (2)$$

There is a unique Nash equilibrium of this game. The total production at this equilibrium is noted $q^N(k_{ISO},k_1)$. If $k_{ISO}<k^{**}$, the capacity of the ISO is binding and the firm generates output. Otherwise the price equals c and the firm does not produce. The production game is represented on figure 1.

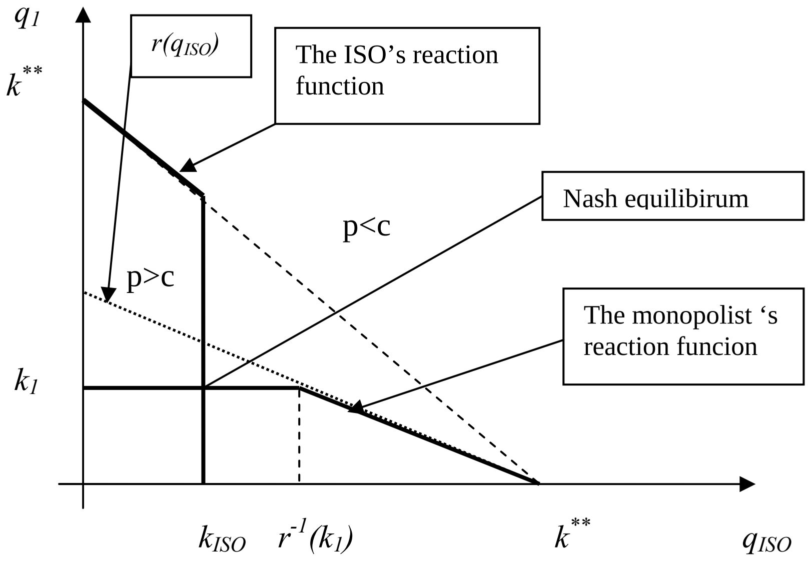


Figure1 : The production stage

The Nash equilibrium of the production subgame is the intersection of the reaction functions. The ISO's reaction function is similar to the straight line p=c for $q_{ISO}<k_{ISO}$. The reaction function of the monopolist is similar to $r(q_{ISO})$ for $r(q_{ISO})<k_1$.

2.2. ISO's choice of capacity

Once the monopolist has chosen his capacity k_1, the ISO has to choose its capacity k_{ISO}. We analyse the best response of the ISO to a capacity k_1. We note $k_{ISO}^+(k_1)$ a strategy of the ISO that maximizes the social surplus W for a given k_1. This strategy might not be unique, if for one k_1 there are several possible choices for the ISO. The ISO maximizes the social surplus subject to an equilibrium constraint, i.e. the production is the unique Nash equilibrium of the production subgame. k_{ISO}^+ should maximizes $W(q^N(k_{ISO},k_1),k_{ISO}+k_1)$. Hence we have :

$$k_{ISO}^+(k_1) \in \arg\max_{k_{ISO}} \int_0^{q^N} p(u)du - c.q^N - I(k_{ISO}+k_1)$$

It is clear that the choice of k_{ISO} could be restricted to the set $\left[0,k^{**}\right]$. In that case the capacity constraint of the ISO is binding at the production stage. Therefore the production is $q^N(k_{ISO},k_1) = k_{ISO} + q_1(k_{ISO},k_1)$. The ISO should consider the influence of its capacity choice on the system's production level. The monopolist's production is continuous and decreasing with respect to the ISO's capacity: it is constantly equal to k_1 for $k_{ISO}<r^{-1}(k_1)$, and equal to $r(k_{ISO})$ for greater k_{ISO}. There is a point of non-derivability at $k_{ISO} = r^{-1}(k_{ISO})$, quantity from which the monopolist's capacity is no longer binding. The total production q^N is increasing with respect to the ISO's capacity, for $-1<r'$.

The derivative of the social surplus with respect to k_{ISO} is:

$$\frac{dW}{dk_{ISO}} = (p-c)\frac{\partial q^N}{\partial k_{ISO}} - I = \begin{cases} (p-c)-I \ \ if \ \ 0<k_{ISO}<r^{-1}(k_{ISO}) \\ (p-c)(1+r')-I \ \ if \ \ r^{-1}(k_{ISO})<k_{ISO}<k^{**} \end{cases}$$

If the ISO chooses a small amount of capacity the firm will bind its capacity at the production stage and it will not use its short-term market power. In this case the social surplus evolution is an usual result: a change in k_{ISO} leads to a similar change in q^N and the social surplus is increased by $p-(c+I)$. However, for k_{ISO} greater than $r^{-1}(k_1)$, the monopolist withholds capacity and his production decreases with respect to k_{ISO}. The impact of this decrease in production on social welfare is represented by the term $(p-c)r'$.

The long-term social optimum could be reached if the monopolist capacity is binding for $k_{ISO} = k^* - k_1$. In that case $q^N = k^*$ and $p=c+I$ the which equals the long-term marginal cost. For the social optimum to be reached, that the following must hold: $k^* - k_1 \leq r^{-1}(k_1)$. In some cases there is a threshold noted k_A such that $k^* - k_1 \leq r^{-1}(k_1)$ if and only if $k_1 \leq k_A$. This quantity is the solution to the following equation:

$$r(k^* - k_A) = k_A$$

This equation has a unique solution on the set $\left[0, k^*\right]$ if and only if $r(0) \leq k^*$, i.e. the production of a monopolist with overcapacity is less than the long-term optimal capacity[3]. This can be easily characterized using the price elasticity ε of the demand at k^*. As the monopolist's profit is strictly quasi concave, $r(0) \leq k^*$ if and only if $p(k^*) + p'(k^*)k^* \leq c$ i.e. $p'(k^*)k^* \leq -I$ [4]. As the price elasticity is defined by: $\varepsilon = \frac{p}{p'.k^*} = \frac{c+I}{p'(k^*).k^*}$, the threshold k_A exists if and only if $-\varepsilon \leq \frac{c+I}{I}$. In most of the following, this condition is assumed to hold. The threshold could be expressed using the elasticity:

$$k_A = -\varepsilon \frac{I}{c+I} k^* \qquad (3)$$

This threshold could be used to compare the total production, when the monopolist's capacity is not binding, with the long-term social optimum:

$$\forall k \in \left[0, k^{**}\right], k \leq k^* - k_A \Leftrightarrow k + r(k) \leq k^*$$

We are now able to establish the following lemma which characterized the best response of the ISO.

Lemma 1:
If $k_1 \leq k_A$:
The social surplus is maximized for $k_{ISO}{}^+(k_1) = k^* - k_1$
The production and the equilibrium price are : $q^N = k^*, p = c+I$
Profits are nil: $\pi_1 = 0, \pi_{ISO} = 0$

If $k_1 > k_A$:
The surplus is maximized for $k_{ISO}{}^+(k_1) \in \left[r^{-1}(k_1), k^* - k_A\right[$
The production and price are $q^N = r(k_{ISO}{}^+) + k_{ISO}{}^+ < k^*, p > c+I$
The ISO's profit is strictly positive: $\pi_{ISO} > 0$.
In this case, the ISO's choice is not unique in general : several values could maximize the social surplus.

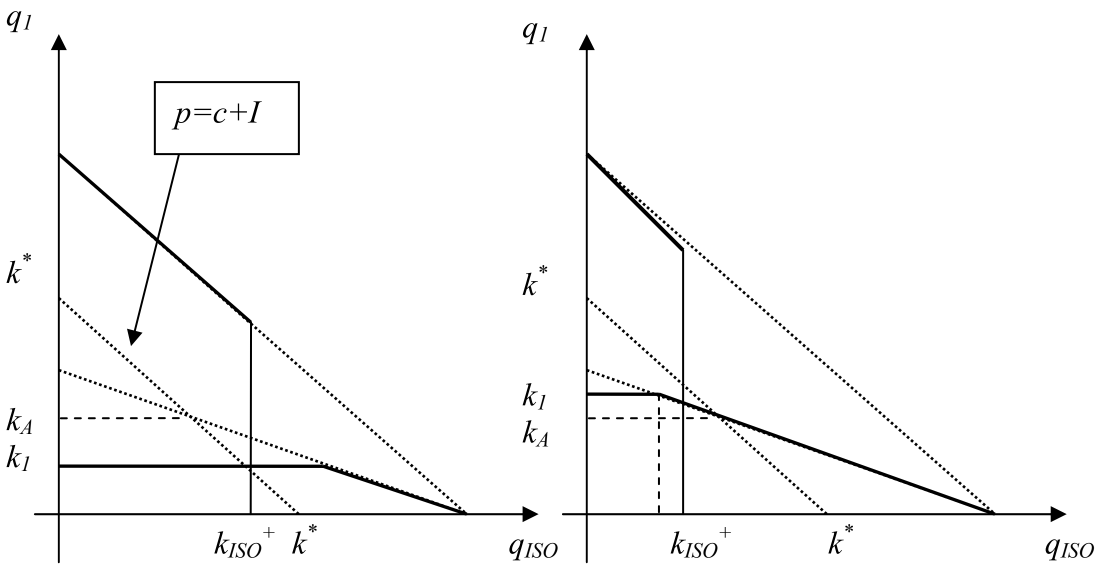
Proof.
The result is due to the following monotonicity properties of the social surplus.
$W(q^N(k_{ISO}, k_1), k_1 + k_{ISO})$ *is increasing with respect to* k_{ISO} *if the capacity of the monopolist is binding at the production stage :* $q_1 = k_1$, *and the total production is less than the long-term optimum:* $q^N = k_1 + k_{ISO} \leq k^*$.
W is decreasing if the production is greater than the long-term optimum: $q^N(k_{ISO}, k_1) \geq k^*$. *For intermediary situations the monotonicity of W is not clear.*

If $k_1 \le k_A$*, the monopolist's capacity is binding for* $k_{ISO} = k^* - k_1$ *. So W is maximized at* $k_{ISO}{}^+ = k^* - k_1$.

If $k_1 > k_A$*, the social surplus is increasing for* $k_{ISO} \le r^{-1}(k_1)$ *and decreasing for* $q^N \ge k^*$ *i.e.* $k_{ISO} \ge k^* - k_A$*. Then W is maximized for* k_{ISO} *in the set* $\left[r^{-1}(k_1), k^* - k_A \right[$ .Q.E.D.

Comments :

The two situations are depicted in figure 2. In the first case (cf figure 2a), the system operator can restore the long-term optimum by completing the total capacity of the industry. In the short-term, the monopolist will produce at full capacity and the price is set equal to the long-term marginal cost. In the second case (cf figure 2b) the short-term market power of the firm prevents the ISO from restoring the long-term optimum. As the ISO considers the loss of social surplus due to new capacity expansions, its incentive to invest is limited. Even if the system operator can force the price down to the long-term marginal price $p=c+I$, it is not socially optimal. Although the profit of the system operator is strictly positive in that case, the profit of the firm could be negative because of unused capacities. The threshold described by formula (2) is linked to the demand elasticity and the share of the sunk cost I in the total cost $c+I$. The less elastic the demand and the smaller the share of sunk costs, the smaller the threshold. It could be easily explained by the links between the exercise of market power and the price elasticity and variable costs. The incentive for the monopolist to restrict production increases with the inelasticity of the demand because of the impact on prices, and with the marginal cost of production.



2a. Case $k_1<k_A$ **2b. Case $k_1>k_A$**

Figure 2 The ISO's best response $k_{ISO}{}^+$ in the two situations of lemma 1: if $k_1<k_A$, then the long-term optimum can be restored such that $k_{ISO}{}^+=k^*-k_1$. If $k_1>k_A$, it cannot be restored.

2.3. The solution of the first stage: the choice of the monopolist.

At the first stage of the game the monopolist considers the reaction of the ISO when choosing his capacity k_1. As stressed above for a quantity inferior to k_A, the firm's profit is zero. At the equilibrium of the game, the firm chooses a quantity sufficiently high so that the ISO could not restore the long-term optimum. The price is then greater than the long-term average cost and if the firm produces at full capacity its profits are strictly positive. The following proposition describes the equilibrium paths of the game according to the value of the different parameters.

Proposition 1 :

If $-\varepsilon \leq \frac{c+I}{I}$, there exist an infinity of subgame perfect equilibriums such that $k_1 \in \left[0, k^*\right], k_{ISO} = k^* - k_1$ and $p=c+I$ and profits are nil.

Otherwise, k_A exists and for any $k^0 \in \arg\max_{k \geq 0} W(k + r(k), k)$, there exists at least one subgame perfect equilibrium whose equilibrium path is : $k_1 = r(k^0) > k_A$ and $k_{ISO} = k^0 < k^* - k_1$ and $q^N = k_1 + k_{ISO}$.

If $\arg\max_{k \geq 0} W(k + r(k), k)$ is a singleton, there is a unique subgame perfect equilibrium path.

Proof:

The first case is trivial; in that case whatever the monopolist's choice, the ISO is able to reach the long-term optimum.

In the second case, the threshold k_A is well defined.

Let $k^0 \in \arg\max W(k + r(k), k))$.

We use k^0 to partially construct an ISO's best response $k_{ISO}{}^+$ that will ensure an equilibrium path as in the proposition. We do not precisely describe the best response of the ISO as we need only some properties.

For $k_1 \geq r(k^0)$, *k^0 maximizes the social surplus:*

$$\forall k_{ISO} \geq r^{-1}(k_1), W = W(k_{ISO} + r(k_{ISO}), k_{ISO} + k_1) = W(k_{ISO} + r(k_{ISO}), k_{ISO}) - Ik_1$$

Hence we can set that $k_{ISO}{}^+(k_1) = k^0$ *for* $k_1 \in \left[r(k^0), r(0)\right]$.

For $k_1 \in \left[k_A, r(k^0)\right]$ *we only need to know that* $k_{ISO}{}^+(k_1) \in \left[r^{-1}(k_1), k^* - k_A\right]$ *(cf lemma 1).*

Next, we state that $k_1 = r(k^0)$ *is the optimal choice for the monopolist given the preceding properties of the ISO's strategy.*

For $k_1 \in \left[r(k^0), r(0)\right]$, *the firm's profit is decreasing because the ISO chooses* $k_{ISO}{}^+(k_1) = k^0$. *Hence* $k_1 - r(k^0)$ *of the firm capacity is excess.*

For $k_1 \in \left[k_A, r(k^0)\right]$ *we know that the ISO capacity is above $r^{-1}(k_1)$ so* $k^0 \leq r^{-1}(k_1) \leq k_{ISO}{}^+(k_1)$. *The production q^N is increasing with respect to the ISO's capacity so* $k_{ISO}{}^+ + r(k_{ISO}{}^+) \geq k^0 + r(k^0)$ *and the price is reduced:*

$$0 \leq p\left(k_{ISO}{}^+ + r(k_{ISO}{}^+)\right) - (c+I) \leq p\left(k^0 + r(k^0)\right) - (c+I)$$

Finally by multiplying by the production and capacity, we get:

$$\left(p - c\right).q_1(k_{ISO}{}^+, k_1) - I.k_1 \leq \left(p - c\right).r(k^0) - I.r(k^0)$$

The profit of the firm is greater for $k_1 = r(k^0)$ *and the firm's profit is strictly positive for* $k_1 = r(k^0)$. *Q.E.D.*

If the monopolist invests less than $r(k^0)$, the ISO increases its investment and the total quantity produced is increased. So the price decreases and so does the firm's profit. The monopolist is therefore induced to increase his investment. But for greater investment than $r(k^0)$ the firm's choice has no influence on the ISO's one and the firm over-invests. Part of her capacity is unused. Although the ISO's profit is strictly positive at equilibrium, the firm cannot increase its investment to get some of this profit. This is because the firm is unable to credibly commit to a given level of production.

The total investment is superior with the ISO's intervention than without. It can be easily proved by comparing the marginal revenues in both situations. In our game, at equilibrium: $c + I \geq c = p + p' k_1 \geq p + p'\left(k_1 + k_{ISO}\right)$, but in the case of a monopolist the investment would satisfy the first order condition: $p + p' k = c + I$. The total marginal revenue is therefore less in our game which implies that the investment is larger.

The uniqueness of a maximizing argument k^0 implies the uniqueness of a subgame perfect equilibrium path. If several arguments maximize $W(k+r(k),k)$, there are several subgame perfect equilibriums with different paths. It is even possible to construct an equilibrium where the firm does not produce at full capacity. It could arise if the ISO's best response switches from one maximizing argument to another.

2.4. Several states of demand and several technologies

As we mentioned, this model could be easily extented to a case with several discrete states of demand and several technologies if contestability or competitiveness is assumed in lower states of demand. The situation is different if we consider that the firm has market power in several states of demand.

If the demand variations are sufficiently important, the social optimum cannot be reached and there is no result similar to lemma 1. It is easily understandable, given that the social optimum could be reached if the firm produces at full capacity (her market power does not limit the ISO's ability to restore the optimum). When demand varies considerably, it is socially optimal not to produce at full capacity in each state of demand. Therefore the firm may use its market power in low states of demand which prevents the ISO from restoring the long-term optimum.

Even if the action of the ISO is limited to the peak period and its technology to peak technology, there is an impact on investment in base-load technology. The result might be counterintuitive. The monopolist no longer uses his peak capacity to equalise his marginal revenue in peak period with the long-term marginal cost of a peaker. This marginal revenue is determined by the game with the ISO. The result of this game is that peak price is above long-term marginal cost and independent of base-load capacity choice. So the revenue obtained by an additional capacity of base-load equipment is greater with the ISO's intervention in peak period. Thus, the firm's investment in base-load capacity is greater.

3. A game between an ISO investor and an oligopoly of generators

We now move on to an oligopoly case. There are n firms indexed by $i=1..n$. They choose their capacity $(k_i)_{i=1..n}$ simultaneously at the first stage, followed by the ISO in the second stage. As above the third stage is a production game with fixed capacities, the production of firm i and ISO are q_i and q_{ISO} respectively. We solve the game by backward induction. Hence, we first analyse the production stage followed by the ISO's choice. Finally we derive some results on the "symmetric" subgame perfect equilibriums of the game. "Symmetric" indicates that every firm choose the same capacity: $k_i = k_j, \forall i, j$ at the first stage; the ISO's choice may be different. The second and third stages are analysed in the general case of asymmetric capacities. Even if such situations do not arise along the equilibrium path of a symmetric equilibrium they are useful to analyze deviations.

3.1. The production stage

The reaction of an individual firm is similar to the reaction of the monopolist described by (2). The function *r(.)* described the reaction of a firm when her capacity constraint is not binding. As we want to analyse the reaction of the entire oligopoly to the production choice of the ISO, we introduce a family of functions representing the reaction of an oligopoly. The reaction of an oligopoly to a fixed production of the ISO is the aggregation of individual reactions to this quantity plus the production of rivals. We note *r(.,m)* is the reaction of *m* firms oligopoly when capacity constraints are not binding. Usual results of Cournot oligopoly ensure that these functions are well defined (cf. Vives 2002).

We have the following relation between the reaction of an individual firm and an oligopoly:

$$m.r\left(\frac{m-1}{m}r(q,m)+q\right)=r(q,m),\forall q\in\left[0,k^{**}\right],\forall m\in\mathbb{N}$$

The reaction of the oligopoly is increasing with respect to the number of firms *m*, and the derivatives with respect to *q* is:

$$\frac{\partial r(q,m)}{\partial q}=\frac{m.r'}{1-(m-1)r'}\in\left]-1,0\right[,\forall m\in\mathbb{N}$$

There exists a unique Nash equilibrium at the production game. The quantities produced at equilibrium depend on the capacities. The production of firm *i* at equilibrium is noted as $q_i^N(k_{ISO},k_1,..,k_n)$. For $k_{ISO}\in\left[0,k^{**}\right]$, the system operator produces at full capacity :

$$q^N(k_{ISO},k_1,..,k_n)=k_{ISO}+\sum_i q_i^N(k_{ISO},k_1,..,k_n)$$

The firms' constraints are successively relaxed when the production of the ISO increases. Constraints of firms with greatest capacities are relaxed first. We formalized in the following lemma.

Lemma 2 :

For an n-tuple $\left(k_i\right)_{i=1..n}$ and k_{ISO}, productions at equilibrium verify:

$$\forall i,j,k_i\geq k_j,\text{ then }q_i^N=k_i\Rightarrow q_j^N=k_j$$

Proof :

Let $i,j,k_i\geq k_j$, *we assume that* $q_i^N=k_i$. *The production faced by firm i (resp. j) is noted* q_{-i}^N *(resp.* q_{-j}^N*), it is the sum of the other firms production and the ISO's one.*

First, $q_{-j}^N=q_{-i}^N+q_i^N-q_j^N=q_{-i}^N+k_i-q_j^N\geq q_{-i}^N$.

So $r(q_{-j}^N)\geq r(q_{-i}^N)\geq k_i$ *for* $q_i^N=k_i$. *Hence,* $r(q_{-j}^N)\geq k_j$ *which implies* $q_j^N=k_j$. *Q.E.D.*

For a fixed n-uple of capacities and k_{ISO}, the total equilibrium production is :

$$q^N=k_{ISO}+k+r(k,m)$$

where *m* is the number of firms whose capacities constraints are not binding, and *k* is the sum of the *n-m* other firms' capacities. According to Lemma 2, the firms whose capacity constraint is binding are the firms with the smallest capacities. So, *k* is the sum of the *n-m* smallest capacities.

3.2. The second stage : the investment choice of the ISO

The system operator is assumed to be benevolent. It maximizes the social welfare *W*. The best response of the ISO to a n-tuple of capacities $\left(k_i\right)_i$ is noted $k_{ISO}^{+}(k_1,..,k_n)$. The ISO maximises *W* subject to equilibrium constraints, and anticipates the third stage equilibrium. To avoid complications due to the existence of several best responses we assume the uniqueness of the ISO's best response.

For a small ISO capacity, all firms produce at full capacity at equilibrium. The first constraint to be relaxed is the constraint of the dominant firm i.e. the firm with the largest capacity. If the dominant firm is firm 1, its constraint is relaxed for k_{ISO} greater than the solution to $k_1 = r\left(k_{ISO} + \sum_{i\neq 1} k_i\right)$ which is $k_{ISO} = r^{-1}(k_1) - \sum_{i\neq 1} k_i$. The social optimum could be reached if this constraint is binding for $k_{ISO} = k^* - \sum_i k_i$. The threshold k_A could be used to establish a result similar to lemma 1.

We restrict our attention to situations where the n-tuple of capacities are in the following set *K*:

$$K = \left\{ (k_i)_{i=1..n} \middle/ \forall i, k_i \leq r\left(\sum_{j\neq i} k_j\right), \sum_i k_i \leq k^* \right\}.$$

The restriction to this set allows us to simplify the statement and the proof of the following lemma. The choices of firm's capacities will be in this set at equilibriums which are considered below.

Lemma 3 :
For $(k_i)_{i=1..n} \in K$
If $\max(k_i, i=1..n) \leq k_A$, the long-term optimum can be reached :

$$k_{ISO}{}^{+} = k^* - \sum_i k_i \;, \forall i, q_i^N = k_i, p = c + I$$

Otherwise *p>c+I.*

Proof.
We assume that $k_1 = \max\left\{k_i, i=1..n\right\}$.

If $k_1 \leq k_A$*, then for* $k_{ISO} = k^* - \sum_i k_i$ *,* $r\left(k_{ISO} + \sum_{i\neq 1} k_i\right) = r\left(k^* - k_1\right) \geq k_1$.

It implies that at the production stag,e the capacity constraint of all firms is binding. The production is $q^N = k^*$ *and W is maximised.*

If $k_1 \geq k_A$*, for* $0 \leq k_{ISO} \leq r^{-1}(k_1) - \sum_{i\neq 1} k_i$ *, all firms produce at full capacity and* $q^N = \sum_i k_i \leq k^*$. *Therefore W is increasing with respect to* k_{ISO}*. For greater* k_{ISO}*, the derivative of W is* $\frac{dW}{dk_{ISO}} = (p-c)\frac{\partial q^N}{\partial k_{ISO}} - I < (p-c) - I$ *which is negative if* $q^N = k^*$*. Q.E.D.*

This result is similar to lemma 1. One should notice that the threshold does not depend on the number of firms. This threshold should be compared with the dominant firm's capacity and not with total capacity. It is due to the fact that the first constraint to be relaxed is the dominant firm's one. The optimum could be reached if this constraint is still binding for a total production of k^*. Otherwise the short-term market power of the oligopolist prevents the ISO from reaching the long-term optimum.

3.3. The first stage: the choice of capacities $(k_i)_i$ of the oligopoly.

As mentioned earlier, we analyse symmetric equilibriums. To construct equilibria similar to the monopoly case, we define quantities $k^0(n)$:

$$k^0(n) = \arg\max_k W(k + r(k,n), k)$$

First, we established that the restriction to the set K is legitimate and that firms produce at full capacity along any subgame perfect equilibrium path.

Lemma 4:

At any subgame perfect equilibrium $(k_i)_{i=1..n} \in K$ and the firms produce at full capacity : $\forall i, q_i^N = k_i$.

Proof.

The assumption of uniqueness of the ISO's best response is the key assumption. We first state that firms produce at full capacity.

If some firms do not produce at full capacity, we know that the dominant firm is among one of them. In that case a slight change of the dominant firm capacity does not influence the ISO's choice because of uniqueness of the ISO's best response. Such a change would increase profits of the dominant firm by diminishing its investment in unused capacities. The deviation is then profitable.

It is now straightforward to establish the result. The set K is defined by 2 conditions:

(i) $\forall i, k_i \le r\left(\sum_{j \ne i} k_j\right)$ *is necessary so that firms produce at full capacity,*

(ii) $\sum_i k_i = k^*$ *is necessary, so that firms get positive profits. Q.E.D.*

Therefore, lemma 3 concerning the threshold can be used to analyse an equilibrium, Although the threshold does not depend on the number of firms in the oligopoly, the two quantities are linked via the set K. When the firms are numerous any symmetric n-tuple in the set K satisfies the condition of the lemma 3. The critical number of firms is determined by the price elasticity of the demand at k^* and the share of the irreversible cost in the total cost. These results constitute the following lemma.

Lemma 5:

The condition $-\varepsilon < \frac{1}{n}\frac{c+I}{I}$ is equivalent to $r(0,n) \le k^*$. And if it is not satisfied, the symmetric elements of K are in the set $[0, k_A]^n$:

$$-\varepsilon \ge \frac{1}{n}\frac{c+I}{I} \Rightarrow \forall (k,..,k) \in K, k \le k_A$$

Proof

By definition $n.r\left(\frac{n-1}{n} r(0,n)\right) = r(0,n)$.

As r(.) is decreasing $r(0,n) < k^* \Leftrightarrow r\left(\frac{n-1}{n}k^*\right) < \frac{k^*}{n}$ *and using the strict quasi concavity of the profit :* $r\left(\frac{n-1}{n}k^*\right) < \frac{k^*}{n} \Leftrightarrow p(k^*) + p'(k^*)\frac{k^*}{n} < c$ *which is equivalent to* $-\varepsilon < \frac{1}{n}\frac{c+I}{I}$ *. The equivalence is proved.*

If $r(0,n) \ge k^*$, $r\left(k^* - \frac{k^*}{n}\right) = r\left(\frac{n-1}{n}k^*\right) \ge \frac{k^*}{n}$ *as mentioned above. So* $\frac{k^*}{n} \le k_A$ *which implies the result. Q.E.D.*

These lemmas pave the way to the following generalization of proposition 1.

Proposition 2

If $-\varepsilon < \frac{1}{n}\frac{c+I}{I}$ then there is a subgame perfect equilibrium such that along the equilibrium path $k_i = \frac{r(k^0(n),n)}{n}, i = 1..n$ and $k_{ISO} = k^0(n)$, *p>c+I* and $\pi_i > 0, i = 1..n$

Otherwise, the set of the paths of symmetric subgame perfect equilibriums is :

$$\left\{ \left(k_i = k\right)_{i=1..n}, k_{ISO} = k^* - nk \middle/ k \in \left[\frac{k^* - k_A}{n-1}, \frac{k^*}{n}\right] \right\}$$

and *p=c+I* and $\pi_i = 0, i = 1..n$.

Proof

If $-\varepsilon < \frac{1}{n}\frac{c+I}{I}$ *then* $r(0,n) < k^*$ *(lemma 5).*

Hence $k^0(n) + r\left(k^0(n),n\right) < k^*$ *and* $\frac{r\left(k^0(n),n\right)}{n} \geq k_A$.

We assume that $k_2 = .. = k_n = \frac{r\left(k^0(n),n\right)}{n}$ *and state that* $k_1 = k_2$ *maximizes firm 1's profit. It depends on the ISO's strategy* $k_{ISO}^+(k_1,..,k_n)$.

It is clear that $k_{ISO}^+(k_2,k_2,.,k_2) = k^0(n)$.

- *For* $k_1 \geq k_2$*:*

The quantity produced by the oligopoly with respect to k_{ISO} *is noted q(.):*

$$q(k_{ISO}) = \begin{cases} k_1 + (n-1)k_2 \text{ if } 0 \leq \mathrm{k_{ISO}} \leq r^{-1}\left(k_1\right) - (n-1)k_2 \\ (n-1)k_2 + r\left((n-1)k_2 + k_{ISO}\right) \text{ if } r^{-1}\left(k_1\right) - (n-1)k_2 \leq \mathrm{k_{ISO}} \leq k^0(n) \\ r\left(k_{ISO},n\right) \text{ if } k^0(n) \leq \mathrm{k_{ISO}} \end{cases}$$

Firm 1's capacity constraint is relaxed first. It is relaxed for $k_{ISO} = r^{-1}(k_1) - (n-1)k_2$ *which is smaller than* $k^0(n)$. *The constraints of the other firms are relaxed for* $k_{ISO} \geq k^0(n)$.

The production of a constrained oligopoly is less than the production of an unconstrained one : $\forall k_{ISO}, q(k_{ISO}) \leq r\left(k_{ISO},n\right)$. *The social surplus is increasing with respect to quantity for p>c so :*

$$W\left(q(k_{ISO}), k_{ISO} + \sum_i k_i\right) \leq W\left(k_{ISO} + r(k_{ISO},n), k_{ISO} + \sum_i k_i\right)$$

And the second term is maximized when $k_{ISO}=k^0(n)$. *Thus,* $k_{ISO}^+ = k^0(n)$ *and the firm 1 over-invests.*

- *For* $k_1 \leq k_2$*:*

The social surplus is increasing as long as all constraints are binding.

The ISO's best response satisfies $k_{ISO}^+ \geq r^{-1}(k_2) - (n-2)k_2 - k_1$ *and the production is greater than* $k^0(n) + r\left(k^0(n),n\right)$. *The price is therefore less than* $p\left(k^0(n) + r\left(k^0(n),n\right)\right)$ *obtained for* $k_1 = k_2$ *and the profit of firm 1 is less than the profit obtained for* $k_1 = k_2$.

Hence $k_1 = k_2$ *is the best response of firm 1.*

If $-\varepsilon < \frac{1}{n}\frac{c+I}{I}$*, then all symmetric subgame perfect equilibriums are cases of lemma 3 . The set describes all the paths.*
For $k_1 = .. = k_n \in K$ *at equilibrium, a profitable deviation for firm 1 should be for a quantity greater than* k_A *and less than* $k^* - (n-1)k_2$ *which is possible if and only if* $k_2 \leq \frac{k^* - k_A}{n-1}$*. Q.E.D.*

Proposition 2 generalises proposition 1. The proof is longer because of the complexity of the short-term reaction of the oligopoly. However, the logic is the same. If an individual firm deviates from a symmetric equilibrium by increasing its capacity, this does not modify the ISO's choice and profits remain unchanged. If a firm deviates by decreasing its capacity, the ISO must increase capacity and consequently, both price and profits decrease. The existence of a suboptimal symmetric equilibrium depends on the number of firms, the price elasticity and the share of irreversible cost in total cost (*I/c+I*). Such an equilibrium exists if and only if $-\varepsilon < \frac{1}{n}\frac{c+I}{I}$. The less elastic is the demand the more numerous should be the firms so that the ISO can restore the long-term social optimum. A decrease of elasticity increases the incentive for firms to limit production on the short-term and the difficulty for the ISO to restore the long-term optimum. Similarly an increase in the variable cost decreases the short-term production. The less the proportion of sunk costs, the more numerous should be the firms for the long-term optimum to be restored at equilibrium.

3.4. Influence of the level of concentration of the oligopoly

A comparative static analysis is difficult in the general case. The monotonicity of the total investment with respect to the number of firms could not be easily deduced with our assumptions. When the concentration is sufficiently low, any symmetric equilibrium leads to the long-term optimum. Yet for a high concentration i.e. low number of firms, the total investment and production are $k^0(n) + r\left(k^0(n), n\right)$, for which the evolution is not evident. Either $k^0(n) = 0$ or it respects the first order condition : $(p-c)\left(1+\frac{\partial r}{\partial q}\right) = I$. Hence its monotonicity depends on $\frac{\partial^2 r}{\partial q \partial n}$ which could be assumed to be negative[5]. It then seems possible that the total investment decreases with respect to the number of firm for high concentrations. It is the case with a linear demand as shown in the following example.

In the linear case :
The demand function is $p(q) = a - bq$.
The reaction function of an 'unconstrained' oligopoly is:

$$r(q,n) = \frac{n}{n+1}\left(\frac{a-c}{b} - q\right)$$

The capacities chosen along a symmetric equilibrium path are:
If $n+1 \leq \sqrt{\frac{a-c}{I}}$, $k_1 = .. = k_n = \frac{1}{n} r\left(k^0(n), n\right) = \frac{(n+1)I}{b}, k^0(n) = \frac{1}{b}\left(a - c - (n+1)^2 I\right)$,
if $\sqrt{\frac{a-c}{I}} \leq n+1 \leq \frac{a-c}{I}, k_1 = .. = k_n = \frac{1}{n+1}\frac{a-c}{b}, k^0(n) = 0$,
and if $\frac{a-c}{I} \leq n+1$, the total investment and production are k^*.
The total investment with respect to *n* is represented in Figure 3.

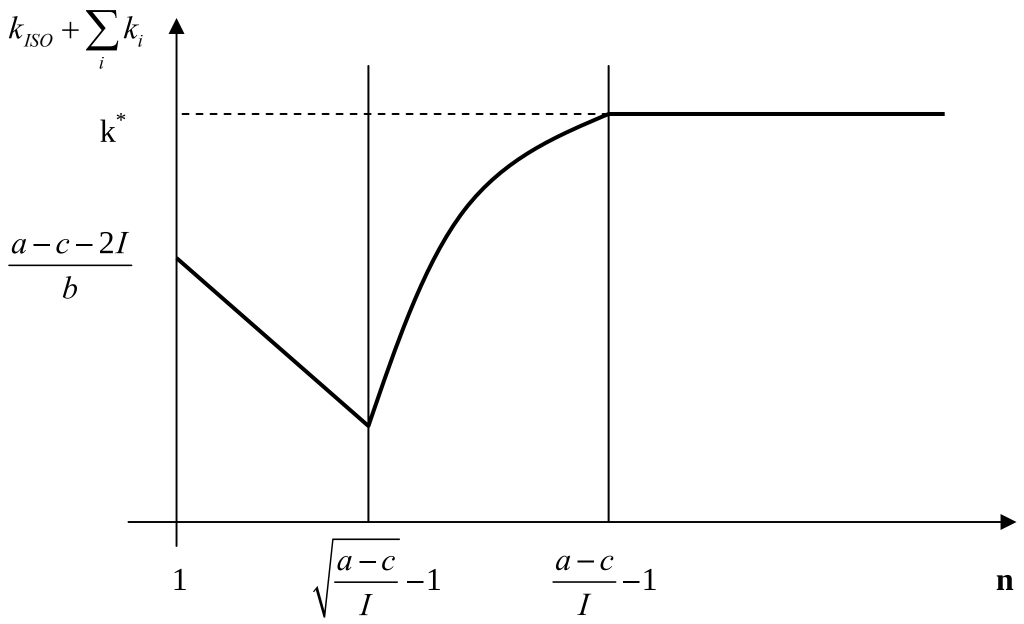


Figure 3 : Total capacity and production with respect to the concentration for a linear demand.

We focused on symmetric equilibriums because of their analytical tractability. Asymmetric equilibria should exist and firms may be able to get strictly positive profits along such equilibria even in the second case of proposition 2.

4. Conclusion

Given the oligopolistic structure of electricity markets, there is a worry that firms might use their long-term market power and under-invest in generation capacity. We have tested here the efficiency of ISO's procurement to compensate this under-investment. In a conventional model of Cournot oligopoly with decision of capacity investment and decision of production joined, such a solution is efficient. The ISO can re-establish the social optimum by completing the investment of the firms. In our model, we dissociate the step of investment and the step of generation. In this setting, the short-term market power changes the social efficiency of the instrument.

We establish that this market power impedes the ISO/regulator to re-establish the social optimum under most of the situations. As he anticipates the influence of his capacity choice on the short-term total production, the increase of the social surplus due to the supplement of production is partly compensated by the decrease of the production of the oligopoly, induced by the exerting of market power. In the long-term the firms remain able to realise strictly positive profits and to keep the sector in a sub-optimal position. However, it may be that this situation is better than the case without a regulator/ISO intervention. Here, the degree of demand elasticity and the level of concentration determine the possibility to move nearer the optimum. For elastic demand functions and (or) decentralised sectors, the optimum could be reached. On the other hand we demonstrate that, if the concentration is high, increasing the number of firms could imply greater difficulty for the regulator/ISO to move closer to the optimum.

An extension of the exercise would be the modelling of a game in which the instrument to circumvent long-term market power is more realistically defined in reference to some experiments. For example, it could allow for the regulator/ISO to develop new equipment ("peakers" for the peak supply and reserve capacity) and produce "out of dispatching" at a zero price. The ISO could recover their costs by a tax or an uplift on all electrical energy output sold by the oligopoly. The main issue is still to understand how the short-term market power could compensate the positive effects of the regulator effects.

References

Boom A.. *Investments in electricity generation capacity under different market structures with price responsive demand*. Diskussionsbeitr¨age des Fachbereichs Wirtschaftswissenschaft 2002/18, Freie Univesit¨at Berlin, September 2002.

Boom A.. *Investments in electricity generating capacity under different market structures and with endogenously fixed demand*. Discussion paper SP II 2003– 01, Wissenschaftszentrum Berlin, June 2003.

Borenstein, 2002 , "The trouble with electricity markets : understanding California's restructuring disaster » Journal of Economic Perspectives

Creti A. et Fabra N., 2003, *Capacity markets for electricity*, Communication to the Colloquium "Efficiency on Electricity Markets", Toulouse, December 2003.

De Vries L.J., 2004, *Securing the public interest in electricity generation markets: the myths of the invisible hand and the copper plate*, PhD thesis, Technical Delft University.

Finon D., 2006, Incentives to invest in liberalisedelectricity industries in North and South. Differences in the need for suitable institutional arrangements, Energy Policy, Vol 24, p.601-618.

Green R., 2001, "Markets for electricity in Europe", *Oxford Review of Economic Policy*, vol 17, n°3.

Joskow P. et Tirole J., 2004, *Reliability and Competitive Markets*, Working paper CEEPR-MIT

Joskow P , 2006*, Competitive electricity markets and investments in new generating capacity*, MIT CEEPR, Working pappoer 06-009, April.

Murphy F.H. and Smeers Y.. *Generation capacity expansion in imperfectly competitive restructured electricity markets*. CORE discussion paper 2002/69, Université Catholique de Louvain, 2002.

Neuhoff K. et De Vries L., 2004, *Investment Incentives for investment in electricity generation,* Utilities Policy, vol. 12(4), p. 253-267, December.

Oren S., 2003, Ensuring generation adequacy in competitive electricity markets, Working Paper UCEI, EPE 007,

Oren S. 2005, *Generation adequacy via call options: safe passage to promised land*, University of California Energy Institute, Working Paper EPE 016, September.

Stoft S., 2002, *Power System Economics, Designing Market for Electricity*. New York, Wiley IEEE

Stoft S; & Crampton , 2006, *The convergence of market designs for adequate generating capapcities (with special attention to the CAISO's resource adequacy problem*, White paper for the electricity Oversight Board;

Turvey R., 2003, *Ensuring adequate generation capacity,* Utilities Policy, Vol. 11, p. 95-102

Perez-Ariaga, et al.(2001), A market approach to long-term security of supply, IEEE transactions on Power Systems, Vol. 25, n°2, p. 349-357.

Spulber, D.F., 1989, *'Markets and regulation'*, Cambridge (Mass.), MIT Press

Vives, X.,2001, *Oligopoly Pricing: old ideas and new tolls*, Cambridge, Mass, MIT Press,

Von der Fehr N.H. and Harbord D.C., 1997*, Capacity investment and competition in decentralised electricity markets*, Department of Economics, (WP n° 27), University of Oslo

NOTES

[1] The log concavity of the demand function is a common assumption. As the function *log p* is concave, the function *p* satisfies $p''p-(p')^2 \leq 0$ on its definition set. Then if *x* is a solution to the first order equation: $p'(q+x).x+p(q+x)=c$ for $q \in \left[0,k^{**}\right]$, substituting the expression of p in the above inequality we get that $p'+p''x \leq 0$, i.e. the cross derivative of the profit is negative, and, $2p'+p''x \leq 0$ the second order condition is satisfied. Hence, the profit function is strictly quasi concave. It ensures the existence and uniqueness of the best response of the monopolist.

[2] The slope of the function *r* is given by: $\frac{p'+p''r}{2p'+p''r}$ which is clearly in the set $]-1,0[$ from the preceding note.

[3] As the slope of *r'* is strictly between 0 and –1 and if the solution of the equation $r(k^*-k)=k$ exists, it is unique. As $r(k^*)>0$, the solution is in the set $\left[0,k^*\right]$ if and only if $r(0)<k^*$.

[4] The strict quasi concavity of the short term profit function $\pi\left(x,y\right)=p(x+y)x-cx$ implies that:

$$\forall z, z \leq r(y) \Leftrightarrow p(z+y)+p'(z+y).z-c=\frac{\partial \pi}{\partial z} \geq 0 .$$

[5] The total investment is $k^0(n)+r\left(k^0(n),n\right)$, if $k^0(n)$ is increasing with respect to *n*, the total investment is also increasing for: $1+\frac{\partial r}{\partial q}>0$ and $\frac{\partial r}{\partial n}>0$. So we wanted to know wether $k^0(n)$ could be decreasing with respect to *n*, by derivating the first ordder condition we get that:

$$k^0\,'(n)=-\frac{p'\frac{\partial r}{\partial n}\left(1+\frac{\partial r}{\partial q}\right)+p\frac{\partial^2 r}{\partial q \partial n}}{p'\left(1+\frac{\partial r}{\partial q}\right)^2+p\frac{\partial^2 r}{\partial q^2}}$$

As r could be assumed to be concave with respect to q, for $k^0(n)$ to be decreasing one requires the crossderivative to be negative. It seems to be a reasonable assumption as $\lim \frac{\partial r}{\partial q}=-1$ when *n* grows to infinity.