OPERATIONS RESEARCH
Vol. 53, No. 4, July–August 2005, pp. 646–661
ISSN 0030-364X | EISSN 1526-5463 | 05 | 5304 | 0646
DOI 10.1287/opre.1050.0211

Generation Capacity Expansion in Imperfectly Competitive Restructured Electricity Markets

Frederic H. Murphy
Fox School of Business and Management, Temple University, Philadelphia, Pennsylvania 19122, fmurphy@temple.edu

Yves Smeers
Department of Mathematical Engineering and Center for Operations Research and Econometrics, Université Catholique de Louvain, Voie du Roman Pays 34, 1348 Louvain-la-Neuve, Belgium, smeers@core.ucl.ac.be

We consider three models of investments in generation capacity in restructured electricity systems that differ with respect to their underlying economic assumptions. The first model assumes a perfect, competitive equilibrium. It is very similar to the traditional capacity expansion models even if its economic interpretation is different. The second model (open-loop Cournot game) extends the Cournot model to include investments in new generation capacities. This model can be interpreted as describing investments in an oligopolistic market where capacity is simultaneously built and sold in long-term contracts when there is no spot market. The third model (closed-loop Cournot game) separates the investment and sales decision with investment in the first stage and sales in the second stage—that is, a spot market. This two-stage game corresponds to investments in merchant plants where the first-stage equilibrium problem is solved subject to equilibrium constraints. We show that despite some important differences, the open- and closed-loop games share many properties. One of the important results is that the prices and quantities produced in the closed-loop game, when the solution exists, fall between the prices and quantities in the open-loop game and the competitive equilibrium.

Subject classifications: electric utilities; existence and characterization of equilibria; noncooperative games; programming; oligopolistic models.
Area of review: Environment, Energy, and Natural Resources.
History: Received June 2001; revisions received June 2002, March 2003, April 2004; accepted April 2004.

1. Introduction: Investments in Power Generation

Capacity expansion models in power generation have evolved into quite complex tools. However, their economics have remained essentially the same as in Massé and Gibrat (1957). Generation plants differ by their investment and operation costs. Capacity expansion models select the mix of plants that minimizes the total cost of satisfying a time-varying demand with randomness over a typical horizon of, say, 20 years. A capacity expansion model designed for a regulated monopoly (e.g., Murphy and Soyster 1983) converts directly into one applicable to a perfectly competitive market. One first introduces a demand model that accounts for the dependence of demand on prices, and then maximizes producers plus consumers surplus to find the equilibrium.

The classic formulation assumes away some important phenomena. Having capacity investments with long lives implies risks. Except for the prudence reviews that developed in the United States (see Joskow 1998), these risks were generally passed on to the consumer. This allowed the industry and the modeler to assume away most risk factors, including the uncertainty of future demand and fuel costs. Even though our goal is to look at capacity investments in competitive situations, we retain this simplification on fuel costs.

Perfect competition is a strong assumption when it comes to restructured electricity markets. A more suitable hypothesis is an oligopolistic market, where each generator can influence prices. Representing imperfect competition as done here is much more complex than representing perfect competition.

Market power is an actively researched area in the literature on restructured electricity systems. Several models exist that look at the operations of a market with oligopolistic players when capacities are given (see, for instance, Wei and Smeers 1999 and the surveys by Daxhelet and Smeers 2001, Hobbs 2001). An extensive stream of less formalized literature also treats the subject. In contrast, very little is available when it comes to investment. Chuang et al. (2001) formulate a single-period Cournot model, as in the second model we present, and solve examples of equilibria. Except for this paper, and to the best of our knowledge, neither qualitative nor quantitative results from market models dealing with both investments and operations in an oligopolistic electricity market exist at this time. However, economic theory does provide several frameworks for looking at the issue.

We begin with strategic investments, which are investments that are made to modify a rival's actions. They are best interpreted in a two-stage decision context where investment decisions are made first, while operations (generation, trading, and sales) are decided in the second stage. Second-stage uncertainties, when they are present, influence first-stage decisions. In the first model of this type, Spence (1977) considers the case where an incumbent builds capacity in the first stage while a potential entrant invests in the second stage. Both operate in the market in the second stage. The potential entrant incurs a fixed cost to enter the market, which the incumbent has already paid. The potential entrant decides to enter the market only if it can make a positive profit after paying for the fixed cost. The incumbent selects its capacity and operating levels to maximize its profit subject to the condition that it wants to bar the entrant from the market. Once the potential entrant decides to not enter, the incumbent operates below its capacity to maximize its profits.

Dixit (1980) retains most elements of Spence's model, but allows the incumbent to add capacity in the second stage. Another difference from Spence is that the second-stage market is Cournot. The problem of the incumbent is thus a mathematical program subject to equilibrium constraints (MPEC) (Luo et al. 1996). MPEC problems are more general than the bilevel mathematical program that subsumes Spence's model.

Gabsewicz and Poddar (1997) assume that the two firms can simultaneously enter a Cournot market with operational decisions made in the second stage. Their model drops the fixed cost to enter the market. Uncertainty is a key element of the Gabsewicz and Poddar (1997) model. They assume that the demand function is revealed in the second stage after the investment is made and that the achieved equilibrium is contingent on this demand information. This nesting of two equilibrium problems (a subgame-perfect equilibrium) leads to a stochastic equilibrium problem subject to equilibrium constraints.

In a different but related context, Allaz (1992) and Allaz and Vila (1993) study the forward commodity markets with market power through an equilibrium model subject to equilibrium constraints. Their models look at the incentives for producers of some commodity to trade in the forward market (first stage) before going into the spot market (second stage). While Allaz (1992) and Allaz and Vila (1993) models do not immediately apply to our investment problem, the former can be adapted to fit a realistic power market by considering two forward electricity markets, namely peak and off peak. Kamat and Oren (2004) have recently extended the work of Allaz and Vila to study some congestion problems arising in restructuring power systems.

The whole field of real options is also relevant to our problem (see Dixit and Pindyck 1994, Trigeorgis 1996 for general presentations, and Ronn 2002 for different applications to the energy field). When applied to investment in generation capacity, this theory describes the value of a new plant by a stochastic process that depends on one or several risk factors (such as prices of electricity and fuels). The investment in a plant is seen as an option that is exerted only when the value of the plant is sufficiently high. Using real options provides an alternative to the treatment of uncertainty for the capacity expansion problem by stochastic programming (Louveaux and Smeers 1988, Janssens de Bisthoven et al. 1988). The extension of real options to market models is more complex. Dixit and Pindyck (1994) provide a treatment of both perfectly and imperfectly competitive markets. Other papers such as those in Grenadier (2000) and the book by Smit and Trigeorgis (2003) report extensions of the approach to game situations. Currently, these latter models do not seem capable of handling the idiosyncrasies of power generation. Finally, the seminal paper by Kreps and Scheinkman (1983) and the subsequent literature provide a framework that could be drawn upon in this context.

One of our objectives is to move a few steps from the economic concepts towards more realistic computable models of capacity expansion in restructured electricity systems. Electricity demand is both time varying and uncertain. The time-varying character of electricity demand is often represented by a load duration curve, the inverse of which can be converted into a probability distribution function. This probabilistic interpretation allows one to incorporate the overall uncertainty underlying demand. That is, some representation of uncertainty, such as that found in Gabszewicz and Poddar (1997), is necessary for dealing with investments in electricity.

Both the oligopolistic investment problem and the issue of entry deterrence are relevant to model the restructured power industry. The oligopolistic problem is more directly applicable to the U.S. situation where investor-owned utilities in restructured systems have largely divested their power plants. In contrast, entry deterrence appears to be directly applicable to the European market where this divestiture process is at an earlier stage and a dominant player remains in place in most countries. The Spence and Dixit models, as well as Schmalensee's (1981) and Bulow et al. (1985) variants, involve fixed costs or economies of scale to deter entrants. There are no scale economies or fixed costs (in the sense of costs independent of installed capacity) associated with the decision to add capacity. Barriers to entry are limited to factors unrelated to these costs. An example of such a barrier is access to sites where capacity can be built. Our representation of the electricity sector incorporates different types of plants to economically satisfy the time-varying demand. We model the oligopolistic investment problems with the players using different technologies having different cost characteristics. This diversity of technologies and agents has consequences. While Gabsewicz and Poddar rely on the symmetry of their problem (both agents use the same technologies) to prove the existence of equilibrium, the asymmetry of our agents can invalidate this existence.

To simplify the problem while retaining key aspects of the power sector, we assume only two types of capacity, namely, peak units and baseload units (Stoft 2002, Chapter 2). Peak units have lower investment and higher operating costs than baseload units. This technological diversity departs from the cited economic literature and is a consequence of electricity not being storable. The load duration curve is discretized into a finite set of demand scenarios. Price responsiveness is included in the form of a price-responsive demand curve in each scenario. One can formulate each player's optimization problem as a two-stage stochastic program with multiple demand curves. Because we assume that operating costs are invariant in our model, the order in which plants are dispatched does not change across multiple demand scenarios and the scenarios collapse into a single scenario. This is very much akin to the Gabsewicz and Poddar (1997) framework.

The paper presents three models, a perfect-competition model, an open-loop Cournot model, and a closed-loop Cournot model (see Fudenberg and Tirole 1991 for these notions). We use perfect competition as a benchmark. The open-loop Cournot model is a relatively simple representation of imperfect competition. Its strategic variables are investment and operations, with these variables selected at the same time. Even though the model is mathematically simple, it has the realistic interpretation that plants are simultaneously built and the output is sold under long-term contract. The model corresponds to an industry structure organized around Power Purchase Agreements (Hunt and Shuttleworth 1996). The distinguishing feature of the closed-loop Cournot model is that capacity decisions are made in the first period and operating decisions in the second period. The closed-loop Cournot model can be seen as an industry structure organized around a spot market (Hunt and Shuttleworth 1996). The generators play against each other when making investments, knowing how they will play against each other when operating their plants. This feature makes the closed-loop game a first-period equilibrium subject to equilibrium constraints in the second period.

Most markets are a mix of spot and contract sales. Allaz (1992) and Kamat and Oren (2004) deal with the question of the division between the two in a market with forward contracts. Our framework does not include forward contracting (e.g., one year), a problem that has received considerable attention in the literature, including in addition to the two mentioned articles Green (1999), Newbery (1998), Wolak (1999), and Bessembinder and Lemmon (2002). The analysis of forward contracting is left to further research.

This paper is organized as follows. The next section (Background) describes the power sector considered in the paper. The equilibrium conditions and standard properties of the perfect-competition case are discussed in §3. The open-loop Cournot model is presented in §4, together with equilibrium conditions and some properties of the solution. A sensitivity analysis of the short-term equilibrium is also presented. These sensitivity results are first applied in §5 to short-term reaction functions and some of their properties. The closed-loop Cournot model is introduced in §6. Its analysis constitutes the core of the rest of the paper. The solutions of the closed- and open-loop Cournot models are compared to establish their similarities and differences. These properties allow one to derive results comparing the investments in both models. Finally, §7 deals with the difficult issue of existence and uniqueness of a solution of the closed-loop Cournot model. Section 8 concludes the paper. To facilitate the exposition, the proofs not presented in the text are in an online appendix. At the end of the paper is a list of all notation used in the paper. Appendix 1 contains the proofs and Appendix 2 provides the background on the competitive model.

2. Background

We consider a simple electricity system where all demand and supply is concentrated at a single node, avoiding network congestion issues. Incorporating congestion is intractable at this stage, and the main forces driving investments in restructured markets are still so unknown that it seems better not to cloud the issue with the impact of congestion. We approximate the load duration curve with a step function (Figure 1).

To simplify notation, we assume that these segments are one unit wide. We index these segments by $s = 1, \ldots, S$, where $s = 1$ is the peak segment and $s = S$ is the base segment.

The different models considered in this paper deal with only two types of generation equipment, each characterized by their annual (per kW) investment (K) and operations (per *kWh*) (ν) cost. We use p to denote a peaker (e.g., gas turbine) and b to denote a base-load plant (e.g., combined cycle gas turbine). By assumption, peakers are cheaper for peak demand, $K_p + \nu_p < K_b + \nu_b$, and base plants are cheaper for base demand, $K_b + S\nu_b < K_p + S\nu_p$. These assumptions are illustrated in Figure 2.

Figure 1. Yearly demand decomposition.

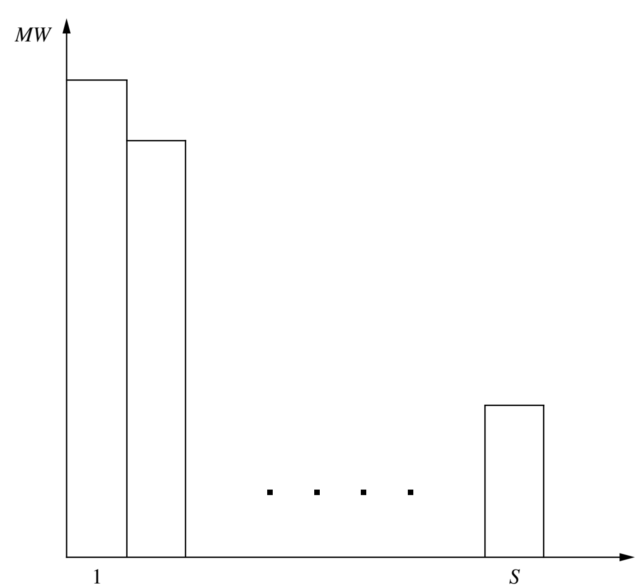
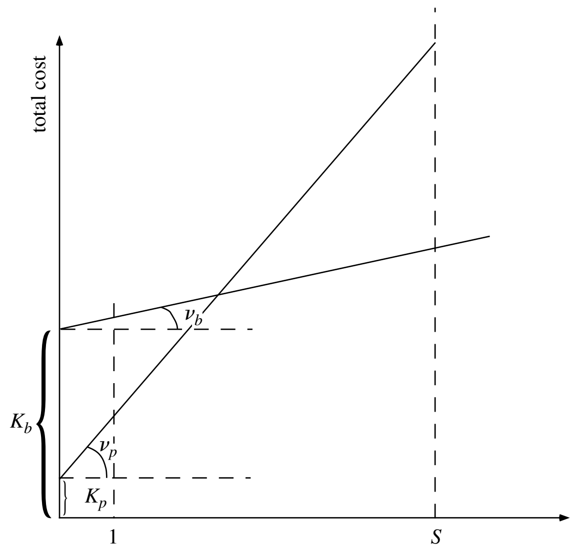


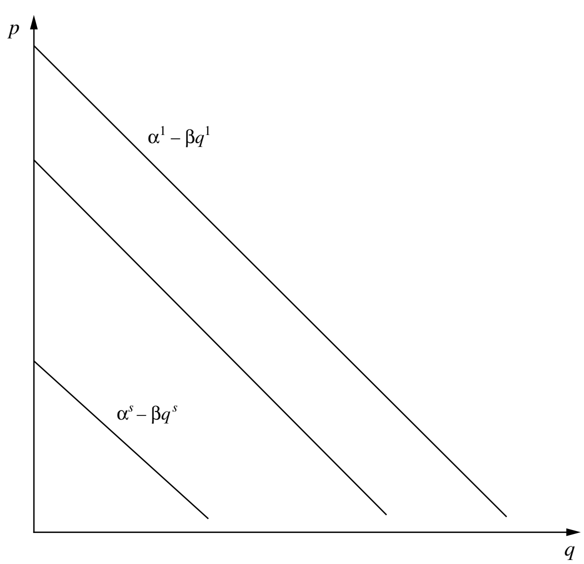
Figure 2. Peak and base plant costs.



Generation capacities are built and operated by two generators denoted $i = p, b$. To simplify the structure of the model, we assume that generator p builds and operates only peak plants, while generator b builds and operates only base plants. Essentially, we are looking at the case where companies develop expertise and specialize. Specialization creates an asymmetry, just as incumbency has for the models in the cited papers. Investment variables are denoted $x_i, i = p, b$ for investments by generators p and b, respectively, and are continuous. Operations variables are denoted $y_i^s, i = p, b; s = 1, \ldots, S$ for the production of generator i in time segment s. Needless to say, we have $x_i \geqslant y_i^s \geqslant 0, i = p, b; s = 1, \ldots, S$.

Finally, demand in each time segment s of the second stage is given by an affine inverted demand curve: $p^s = \alpha^s - \beta q^s, s = 1, \ldots, S$ and $\alpha^s, \beta > 0$. Note that there are no cross elasticities between load segments. Thus, there is no representation of load shifting. We use this demand model for two reasons. First, it is a good approximation to a nonlinear demand curve in the immediate neighborhood of the equilibrium. Second, it makes the mathematics of the proofs simpler and more understandable. We use the same slope for all steps to simplify the notation and some of the resulting formulas. What is critical to the character of our results is that the demand curves do not cross. Demand is higher in the peak segment and decreases as one moves towards the base segments. This is expressed as $\alpha^1 > \alpha^2 > \cdots > \alpha^S$. So that there is positive production, we assume that $\alpha^1 \geqslant \nu_b$. The inverted demand curves for the different time segments are depicted in Figure 3.

Figure 3. Inverted demand curves.



3. The Perfect-Competition Case: Equilibrium Conditions

First, consider the case where both generators compete with given capacities x without exerting market power. That is, they generate until marginal cost equals the market price. Each of the generators has the following optimization problem when it takes the prices p^s as given:

$$\min_{x_i, y_i^s} - \sum_s [p^s - \nu_i] y_i^s + K_i x_i$$
$$\text{s.t.} \quad x_i - y_i^s \geqslant 0, \quad y_i^s \geqslant 0, \quad s = 1, \ldots, S. \tag{1}$$

Let ω_i^s be the dual on the nonnegativity of y_i^s and λ_i^s the dual on the capacity upper bound. Let $\{1, \ldots, S_i\}$ be the load segments for which capacity of player i is binding. Note that outside the context of the equilibrium, because the prices are fixed, the solutions of player i are $x_i = 0$ when $\sum_{s=1}^{S_i} p_s < K_i + S_i \nu_i \ \forall S_i$; $x_i = \infty$ when $\sum_{s=1}^{S_i} p_s > K_i + S_i \nu_i$ for some S_i; and $x_i = [0, \infty)$ when $\sum_{s=1}^{S_i} p_s = K_i + S_i \nu_i$ for all S_i.

Take the Karush-Kuhn-Tucker (KKT) conditions from (1) for both players, using $-i$ to index the player other than i. After replacing p_s by its expression as stated in the inverted demand curve, it can be shown that the vector of generation levels $y_i^s, s = 1, \ldots, S, \ i = p, b$ at equilibrium satisfies the following short-term (operations) equilibrium conditions where the producer does see the demand response to price (see Sherali et al. 1982 for the derivation of these conditions in the fixed-demand case).

$$-\alpha^s + \beta y_i^s + \beta y_{-i}^s + \nu_i + \lambda_i^s = \omega_i^s \geqslant 0, \quad y_i^s \geqslant 0, \ \omega_i^s y_i^s = 0,$$
$$x_i - y_i^s \geqslant 0, \quad \lambda_i^s \geqslant 0, \quad (x_i - y_i^s)\lambda_i^s = 0, \tag{2}$$
$$i = p, b; \quad s = 1, \ldots, S.$$

Equilibrium capacity levels x satisfy the following long-term (investment) equilibrium conditions:

$$K_i - \sum_{s=1}^{S} \lambda_i^s \geqslant 0, \quad x_i \geqslant 0, \ x_i \left(K_i - \sum_{s=1}^{S} \lambda_i^s \right) = 0, \quad i = p, b. \tag{3}$$

Some efficiency properties derived from these equilibrium conditions are given in Appendix 2.

4. The Open-Loop Cournot Model

4.1. Equilibrium Conditions

We now take up the first imperfect-competition model, referred to as the open-loop Cournot model. In this model, each generator selects its capacity x_i and generation plan y_i^s, taking the generation levels y_{-i}^s of the other player as given. In short, generator i, $i=p,b$, solves the continuous quadratic programming problem,

$$\min_{x_i,\, y_i^s} \ \sum_s[-\alpha^s+\beta(y_i^s+y_{-i}^s)]y_i^s+\nu_i\sum_{s=1}^S y_i^s+K_ix_i$$
$$\text{s.t.}\quad x_i-y_i^s\geqslant 0,\quad y_i^s\geqslant 0,\quad s=1,\ldots,S. \tag{4}$$

The solution to this problem satisfies the following short-term equilibrium conditions, which are the KKT conditions for each player:

$$-\alpha^s+2\beta y_i^s+\beta y_{-i}^s+\nu_i+\lambda_i^s=\omega_i^s\geqslant 0,\quad y_i^s\geqslant 0,\quad \omega_i^s y_i^s=0,$$
$$x_i-y_i^s\geqslant 0,\quad \lambda_i^s\geqslant 0,\quad (x_i-y_i^s)\lambda_i^s=0, \tag{5}$$
$$i=p,b;\quad s=1,\ldots,S.$$

As in the perfect-competition case, we obtain the equilibrium conditions by stating that the KKT conditions for both players are satisfied simultaneously. The solution also satisfies the long-term equilibrium conditions (3). Efficiency properties for this equilibrium are discussed in Appendix 2.

4.2. Solution Existence and Uniqueness

We use the theory of variational inequalities (Harker and Pang 1990) to analyze the existence and uniqueness of the above Cournot model. Define for $s=1,\ldots,S$,

$$y^s=\begin{pmatrix} y_p^s \\ y_b^s\end{pmatrix},\quad G_i^s(y^s)\equiv-\alpha^s+2\beta y_i^s+\beta y_{-i}^s+\nu_i,\quad i=p,b. \tag{6}$$

(Note that $-G_i^s(y^s)$ is equal to the marginal revenue minus the short-run marginal cost of generator i.) Also define

$$y=(y^1,\ldots,y^S),\quad G^{sT}(y^s)=(G_p^s(y^s),G_b^s(y^s)),$$
$$G^T(y)=(G^{1T}(y^1),\ldots,G^{ST}(y^S)), \tag{7}$$
$$x=(x_p,x_b),\quad K^T=(K_p,K_b),\quad F^T(x,y)=(K^T,G^T(y)). \tag{8}$$

Let Z be the set of feasible (x,y).

Definition 1. The solution to the variational equality $VI(Z,F)$ is a point

$$z^*=\begin{pmatrix} x^* \\ y^*\end{pmatrix}$$

belonging to Z satisfying

$$F(z^*)^T(z-z^*)\geqslant 0 \tag{9}$$

for all $z\in Z$.

The mapping $F(z)$ is monotone if for all $(x,y)\in Z$, $[F(z^1)-F(z^2)]^T(z^1-z^2)\geqslant 0$. It is strictly monotone when this inequality is strict whenever $z^1\neq z^2$.

The following lemma provides the basic technical result for analyzing the open-loop Cournot model.

Lemma 1. *$G(y)$ is strictly monotone, $F(x,y)$ is monotone.*

It is now possible to restate the open-loop Cournot competition problem as:

Seek (x^*,y^*): $x_i^*-y_i^{s*}\geqslant 0$, $y_i^{s*}\geqslant 0$, $i=p,b$; $s=1,\ldots,S$ satisfying

$$F(x^*,y^*)^T(x-x^*,y-y^*)\geqslant 0 \tag{10}$$

for all (x,y): $x_i-y_i^s\geqslant 0$, $y_i^s\geqslant 0$, $i=p,b$; $s=1,\ldots,S$.

The properties of this model are summarized in the following theorem, which also invokes the notion of dynamic consistency introduced in Newbery (1984). The multiperiod solution of a game is dynamically consistent when the optimal actions for future periods as part of the Period 1 solution remain optimal in the subsequent periods, given the first-period solution. Note that dynamic consistency is a weaker concept than subgame perfection (see Haurie et al. 1999 for a discussion of these two concepts).

Theorem 1. *Given the assumptions made on demand and technology, there always exists an open-loop Cournot equilibrium, and it is unique. This equilibrium is dynamically consistent. The base player always invests a positive amount at equilibrium. The peak player does not necessarily do so, except if the equilibrium demand in some segment $s\leqslant S_p$ is larger than the equilibrium demand in segment S_{p+1}.*

4.3. Sensitivity Analysis

This section presents a set of results that pave the way towards the analysis of the closed-loop equilibrium problem. These results show how the solution of the short-term equilibrium problem varies with the generation capacities and the demand parameters and are fundamental to the properties of both the open-loop and closed-loop games. Using Theorem 1 and the variants introduced below, we can state the following definitions and lemma.

Definition 2. Let $y_i^s(x)$, $i=p,b$; $s=1,\ldots,S$ be the solution of the short-term equilibrium conditions (5) for a given x, the vector of capacity for both players.

The $y_i^s(x)$ satisfy the following properties.

Lemma 2. *$y_i^s(x)$ is well defined (y_i^s is unique for given x). Each $y_i^s(x)$ is left and right differentiable with respect to x_j, $j=i,-i$.*

Now consider the solution of the short-run equilibrium in time segment s as a function of the demand level in that time segment. The solutions satisfy the following properties.

LEMMA 3. *Define $y_i^s(\alpha^s)$, $\lambda_i^s(\alpha^s)$, and $\omega_i^s(\alpha^s)$ to be the solutions of the short-run equilibrium conditions (5) as functions of α^s. $y_i^s(\alpha^s)$ and $\lambda_i^s(\alpha^s)$ are monotonically nondecreasing in α^s. $\omega_i^s(\alpha^s)$ is monotonically decreasing in α^s when nonzero.*

This result is intuitive. It says that the generation level of each agent increases with the willingness to pay (α^s) for electricity. It also states that the marginal value of capacity in some time segment increases with the willingness to pay for electricity in that time segment. Because willingness to pay for electricity in the different time segments decreases with the index of these time segments, this lemma implies the following corollary.

COROLLARY 1. *If $y_i^s = x_i$, then $y_i^{s'} = x_i$ for $s' < s$.*

The peak generator has a higher operating cost than the base generator. Thus, barring the case where base generation is limited by available capacity, peak generation is lower than base generation in any given time segment. This is stated in Lemma 4.

LEMMA 4. *$y_p^s < y_b^s$ in any time segment s such that the baseload capacity is not binding ($y_b^s < x_b$).*

For a given vector x of generation capacities, it is possible to partition the set of the different time segments into the following different classes.

$$
\begin{aligned}
&\text{(a)}\quad -\alpha^s + 2\beta x_i + \beta x_{-i} + \nu_i + \lambda_i^s = 0,\\
&\qquad 0 < y_i^s = x_i, \quad \lambda_i^s \geqslant 0,\ i = p, b.\\
&\text{(b)}\quad -\alpha^s + 2\beta y_i^s + \beta y_{-i}^s + \nu_i = 0,\\
&\qquad 0 < y_i^s < x_i, \quad \lambda_i^s = 0,\ i = p, b.\\
&\text{(c)}\quad -\alpha^s + 2\beta x_i + \beta y_{-i}^s + \nu_i + \lambda_i^s = 0,\\
&\qquad 0 < y_i^s = x_i, \quad \lambda_i^s \geqslant 0,\\
&\quad -\alpha^s + \beta x_i + 2\beta y_{-i}^s + \nu_{-i} = 0, \qquad (11)\\
&\qquad 0 < y_{-i}^s < x_{-i}, \quad \lambda_{-i}^s = 0.\\
&\text{(d)}\quad -\alpha^s + \beta y_{-i}^s + \nu_i = \omega_i^s, \quad y_i^s = 0, \quad \omega_i^s \geqslant 0,\\
&\quad -\alpha^s + 2\beta y_{-i}^s + \nu_{-i} = 0, \quad 0 < y_{-i}^s < x_{-i}, \quad \lambda_{-i}^s = 0.\\
&\text{(e)}\quad -\alpha^s + \beta x_{-i} + \nu_i = \omega_i^s, \quad y_i^s = 0, \quad \omega_i^s \geqslant 0,\\
&\quad -\alpha^s + 2\beta x_{-i} + \nu_{-i} + \lambda_{-i}^s = 0, \quad 0 < y_{-i}^s = x_{-i},\\
&\qquad \lambda_{-i}^s \geqslant 0.
\end{aligned}
$$

Each class has properties that influence the capacity equilibrium. In the rest of the paper we refer to these load segment types by these letters as we develop the properties.

Define $B_i y_j(x) = \partial y_j / \partial x_i$ when the derivative exists. The derivative of the second-stage equilibrium variables with respect to the first-stage capacities can be characterized as follows.

LEMMA 5. *The derivative of y_j with respect to x_i, when it exists, can be stated as follows for the above cases:*

$$
\begin{aligned}
&\text{(a)}\quad B_i y_i(x) = 1, \quad i = p, b, \quad B_i y_{-i}(x) = 0, \quad i = p, b.\\
&\text{(b)}\quad B_i y_j(x) = 0, \quad i = p, b; \quad j = p, b.\\
&\text{(c)}\quad B_i y_i(x) = 1, \quad B_{-i} y_{-i}(x) = 0,\\
&\qquad B_{-i} y_i(x) = 0, \quad B_i y_{-i}(x) = -1/2. \qquad (12)\\
&\text{(d)}\quad B_i y_j(x) = 0, \quad i = p, b; \quad j = p, b.\\
&\text{(e)}\quad B_{-i} y_{-i}(x) = 1, \quad B_i y_i(x) = 0,\\
&\qquad B_i y_{-i}(x) = 0, \quad B_{-i} y_i(x) = 0.
\end{aligned}
$$

5. Reaction Curves

Reaction curves play a significant role in the study of oligopolistic equilibria in economics. To conduct our analysis, consider the following definition of the short-term reaction curve.

DEFINITION 3. The short-run reaction curve of player i with respect to the action y_{-i}^s of player $-i$ in time segment s, for given capacities x, is the solution of the system

$$-\alpha^s + 2\beta y_i^s + \beta y_{-i}^s + \nu_i + \lambda_i^s = \omega_i^s \geqslant 0; \quad y_i^s \geqslant 0, \quad \omega_i^s y_i^s = 0,$$

$$x_i - y_i^s \geqslant 0, \quad \lambda_i^s \geqslant 0, \quad (x_i - y_i^s)\lambda_i^s = 0.$$

It is denoted $y_i^s(y_{-i}^s; x)$.

This reaction curve satisfies the following property.

LEMMA 6. *$y_i^s(y_{-i}^s; x)$ exists and is well defined. It is piecewise affine with*

$$
\begin{aligned}
\frac{dy_i^s}{dy_{-i}^s} &= 0 \qquad \text{when either } y_i^s = x_i \text{ and } \lambda_i^s > 0\\
&\qquad\qquad \text{or } y_i^s = 0 \text{ and } \omega_i^s > 0\\
&= -\tfrac{1}{2} \quad \text{whenever } 0 < y_i^s < x_i.
\end{aligned}
$$

When $y_i^s = x_i$ and $\lambda_i^s = 0$, or $y_i^s = 0$ and $\omega_i^s = 0$, the left and right derivatives are either 0 or $-1/2$ in the obvious directions based on Lemma 5.

6. The Closed-Loop Cournot Model

6.1. Definition and Equilibrium Conditions

To define generator i's problem in the closed-loop Cournot model, consider first the solution $y_i^s(x_i, x_{-i})$ of the short-run equilibrium conditions (5) with x given. The long-run problem of generator i is then

$$\min_{x_i \geqslant 0} K_i x_i + \sum_{s=1}^{S} \big[-\alpha^s + \beta(y_i^s(x_i, x_{-i}) + y_{-i}^s(x_i, x_{-i})) + \nu_i\big] y_i^s(x_i, x_{-i}). \qquad (13)$$

By definition, (x_p^*, x_b^*) is a subgame-perfect equilibrium (Selten 1975) or a closed-loop Cournot equilibrium (Fudenberg and Tirole 1991, Haurie et al. 1999) if x_i solves generator i's long-run problem for given x_{-i}^*, $i = p, b$.

To characterize the solution to this problem, consider a point $x \geqslant 0$ such that the $y_i^s(x)$ are differentiable. A closed-loop equilibrium at such a point satisfies the condition

$$K_i + \sum_s \left[-\alpha_s + 2\beta y_i^s(x) + \beta y_{-i}^s(x) + \nu_i\right] B_i y_i^s(x) + \sum_s \beta y_i^s(x) B_i y_{-i}^s(x) = \xi_i \geqslant 0, \quad \xi_i x_i = 0. \tag{14}$$

We disregard points of nondifferentiability for the immediate discussion and characterize an equilibrium point where all $y_i^s(x)$ are differentiable. This is done by investigating the relationship between the solution of the closed- and open-loop problems. The analysis later in the paper is fully general.

6.2. Closed-Loop Versus Open-Loop Cournot Model

The following lemma is intuitively reasonable: If one player does not generate in some time segments in the Cournot equilibrium, it must be the one with higher short-term operating costs—that is, the peak player.

Lemma 7. *Suppose that the closed-loop Cournot equilibrium problem has a solution with time segments of type e. Then, the peak player is the one operating at zero level in these time segments.*

This result allows one to derive a first characterization of the relation between the closed- and open-loop problems. It gives a sufficient condition for the two equilibria to be identical. Essentially, this happens when neither player has load segments where the operating decisions change in response to the other player's capacity decision.

Theorem 2. *When each segment s in the closed-loop Cournot equilibrium problem is one of the following types, e with $\omega_i^s > 0$, a, b or d, the equilibrium is the same as the solution of the open-loop Cournot problem.*

Proof. Consider a point of differentiability and the associated equilibrium conditions

$$K_i + \sum_s \left[-\alpha^s + 2\beta y_i^s(x) + \beta y_{-i}^s(x) + \nu_i\right] B_i y_i^s(x) + \sum_s \beta y_i^s(x) B_i y_{-i}^s(x) = 0, \quad i = p, b.$$

The $B_i y_{-i}^s(x)$ are zero when s belongs to a, b, or d or e when $\omega_i^s > 0$ (Lemma 5). The equilibrium condition becomes

$$K_i + \sum_s \left[-\alpha^s + 2\beta y_i^s(x) + \beta y_{-i}^s(x) + \nu_i\right] B_i y_i^s(x) = 0.$$

This expression can be rewritten (using Lemma 5) as

$$K_i + \sum_{s \in a \cup e} \left[-\alpha^s + 2\beta y_i^s(x) + \beta y_{-i}^s(x) + \nu_i\right] = K_i - \sum_s \lambda_i^s = 0,$$

which shows that the solution is also a solution to the open-loop Cournot problem. □

The following can be seen as a restatement of this result in terms of the investment criterion. Specifically, the equality between the K_i and $\sum_i \lambda_i^s$ will play an important role in relating the open- and closed-loop equilibria. The two models produce the same equilibria if and only if $K_i = \sum_i \lambda_i^s$.

Corollary 2. *If for every load segment $x_i = y_i^s$ implies $x_{-i} = y_{-i}^s$, then $K_i = \sum_s \lambda_i^s$.*

This corollary states that in this case the solution of the player i optimization subject to the equilibrium constraints is the same as in the pure optimization in the open-loop game. Theorem 2 indicates that differences between the open- and closed-loop equilibria require the solution to have time segments of type c or e with $\omega_i^s = 0$. These are the segments where the capacity decisions of one player affect the operating decisions of the other. A first characterization of a solution with time segments of type c is given by the following lemma, which states that if the solution has multiple segments of type c in the equilibrium, then the same player is below capacity in all of these segments.

Lemma 8. *In case c, $0 < y_i^s = x_i$ and $0 < y_{-i}^s < x_{-i}$ for some segment s implies that there is no segment s' for which $0 < y_{-i}^{s'} = x_{-i}$ and $0 < y_i^{s'} < x_i$.*

This lemma allows us to introduce the following theorem, which establishes a major divergence between the solution of the open- and closed-loop Cournot equilibria. That is, the solution of the player's optimization subject to equilibrium constraints is different from the optimization in the open-loop game, and the KKT conditions are violated.

Theorem 3. *Consider the case in which some segments fall into c and the other segments fall into one of the four cases a, b, d, and e with $\omega_i > 0$. Then, the solution to the closed-loop Cournot equilibrium problem is different from the solution of the open-loop Cournot equilibrium problem. Moreover,*

$$K_i > \sum_{s=1}^{S} \lambda_i^s, \quad K_{-i} = \sum_{s=1}^{S} \lambda_i^s \tag{15}$$

for the i where $y_i^s = x_i$ and $y_{-i}^s < x_i$ in the segments of type c.

Proof. Using the reasoning of the proof of Theorem 2, one can restate the equilibrium condition as

$$K_i + \sum_{s \in a \cup c \cup e} (-\lambda_i^s) + \sum_{s \in c} \beta y_i^s(x) B_i y_{-i}^s(x) = 0, \quad i = p, b.$$

Suppose that $y_i^s = x_i$ and $y_{-i}^s < x_{-i}$ for $i \in c$. Then, $B_i y_{-i}^s(x) = -1/2$. We obtain

$$K_i - \sum_s \lambda_i^s - \tfrac{1}{2} \sum_{s \in c} \beta y_i^s(x) = 0 \quad \text{and} \quad K_{-i} - \sum_s \lambda_{-i}^s = 0 \quad \text{or}$$

$$K_i = \sum_s \lambda_i^s + \tfrac{1}{2} \sum_{s \in c} \beta y_i^s(x) > \sum_s \lambda_i^s \quad \text{and} \quad K_{-i} = \sum_s \lambda_{-i}^s. \quad \square$$

The interpretation of the theorem is as follows. The investment cost of some plant, at the closed-loop equilibrium, may be higher than the sum of its short-term marginal values in the different time segments as normally implied by the KKT conditions. The difference between the two characterizes the value for the player of being able to manipulate the short-term market by its first-stage investments. This value is not captured in the standard single-stage Cournot model.

The following lemma establishes a relatively intuitive property that is common to the solution of the open- and closed-loop equilibria. It states that the solution of the short-run equilibrium first takes advantage of the existing capacity with low operating costs. As expected, this holds both in the open- and closed-loop equilibria.

Lemma 9. *Assume an equilibrium of the open- or closed-loop Cournot equilibrium problem. At such an equilibrium, if the peak player is at capacity in some time segment s, then the base player is also at capacity in that time segment.*

The following corollary takes advantage of Lemma 9 to refine the result expressed by Theorem 3. It says that if closed- and open-loop equilibria differ, the base player manipulates the short-run market through investment. Accordingly, the per-unit investment cost in the base plant is higher than the sum of the marginal values of this plant in the different time segments.

Corollary 3. *If there exists a closed-loop equilibrium with time segments of type c, then*

$$K_b > \sum_{s=1}^{S} \lambda_b^s \quad \text{and} \quad K_p = \sum_{s=1}^{S} \lambda_p^s. \tag{16}$$

The next results complete the comparison between the closed- and open-loop equilibria. Cournot equilibria are known to reduce quantities put on the market. Theorem 4 says that this effect is less pronounced with the closed-loop game.

Theorem 4. *Suppose that there exists a closed-loop equilibrium. Then, the total capacity in the closed-loop equilibrium is at least as large as the total capacity in the open-loop equilibrium and is larger when there are segments of type c or e.*

Proof. Let o and c indicate the closed- and open-loop Cournot equilibrium, respectively. Suppose that $x_p^o + x_b^o > x_p^c + x_b^c$. We prove the contradiction in two parts.

Part 1. We show that $x_p^o > x_p^c$ and $x_p^o + x_b^o > x_p^c + x_b^c$ implies $\sum_s \lambda_p^{sc} > K_p$, which contradicts the above corollary.

Part 2. We show that $x_b^o > x_b^c$ and $x_p^o + x_b^o > x_p^c + x_b^c$ implies $\sum_s \lambda_b^{sc} > K_b$, which again contradicts the above corollary.

Part 1. Suppose that $x_p^o > x_p^c$, $x_p^o + x_b^o > x_p^c + x_b^c$. We know that $K_p = \sum_s \lambda_p^{sc}$ (long-term equilibrium condition (3) of the open-loop problem), and now show that $\lambda_p^{sc} \geqslant \lambda_p^{so}$ for all s with some inequalities holding strictly. Let $K_p = \sum_{s'} \lambda_p^{s'o} + \sum_{s''} \lambda_p^{s''o}$ with $\lambda_p^{s'o} > 0$ and $\lambda_p^{s''o} = 0$,

$$\lambda_p^{s'o} > 0 \quad \text{implies} \quad y_p^{s'o} = x_p^o > x_p^c \geqslant y_p^{s'c}. \tag{17}$$

By Lemma 9, $\lambda_p^{s'o} > 0$, implies $\lambda_b^{s'o} > 0$, and hence $y_b^{s'o} = x_b^o$. Adding up $y_p^{s'o}$ and $y_b^{s'o}$, one gets

$$y_p^{s'o} + y_b^{s'o} = x_p^o + x_b^o > x_p^c + x_b^c \geqslant y_p^{s'c} + y_b^{s'c}. \tag{18}$$

Adding (17) and (18), one gets

$$2y_p^{s'o} + y_b^{s'o} > 2y_p^{s'c} + y_b^{s'c},$$

and hence,

$$\lambda_p^{s'c} > \lambda_p^{s'o}.$$

Therefore,

$$\sum_{s'} \lambda_p^{s'c} + \sum_{s''} \lambda_p^{s''c} > \sum_{s'} \lambda_p^{s'o} + \sum_{s''} \lambda_p^{s''o} = K_p,$$

which is the desired contradiction.

Part 2. Suppose that $x_p^o \leqslant x_p^c$, $x_b^o > x_b^c$, and $x_p^o + x_b^o > x_p^c + x_b^c$. Let $K_b = \sum_{s'} \lambda_b^{s'o} + \sum_{s''} \lambda_b^{s''o}$ with $\lambda_b^{s''o} = 0$ and $\lambda_b^{s'o} > 0$,

$$\lambda_b^{s'o} > 0 \quad \text{implies} \quad y_b^{s'o} = x_b^o > x_b^c \geqslant y_b^{s'c}. \tag{19}$$

When $\lambda_p^{s'o} > 0$,

$$y_p^{s'o} = x_p^o \quad \text{and} \quad y_p^{s'o} + y_b^{s'o} = x_p^o + x_b^o > x_p^c + x_b^c \geqslant y_p^{s'c} + y_b^{s'c}. \tag{20}$$

Adding (19) and (20), one gets

$$y_p^{s'o} + 2y_b^{s'o} > y_p^{s'c} + 2y_b^{s'c},$$

and hence,

$$\lambda_b^{s'c} > \lambda_b^{s'o}.$$

When $\lambda_p^{s'o} = 0$, the short-term equilibrium conditions of player p in the open- and closed-loop games are

$$-\alpha^s + 2\beta y_p^{s'o} + \beta x_b^o + \nu_p = 0$$

and

$$-\alpha^s + 2\beta y_p^{s'c} + \beta y_p^{s'c} + \nu_p + \lambda_p^{s'c} = 0,$$

which gives

$$2\beta y_p^{s'o} + \beta y_b^{s'o} = \alpha^s - \nu_p \geqslant \alpha^s - \nu_p - \lambda_p^{s'c} = 2\beta y_p^{s'c} + \beta y_b^{s'c}. \tag{21}$$

Adding (19) multiplied by 3β to (21) and simplifying, one gets

$$y_p^{s'o} + 2y_b^{s'o} > y_p^{s'c} + 2y_b^{s'c},$$

and hence,

$$\lambda_b^{s'c} > \lambda_b^{s'o}.$$

We then get

$$\sum_{s'} \lambda_b^{s'c} + \sum_{s''} \lambda_b^{s''c} > \sum_{s'} \lambda_b^{s'o} + \sum_{s''} \lambda_b^{s''o} = K_b,$$

which is the desired contradiction. □

The explanation of the above phenomenon can be found in the capability of the base player to manipulate the short-term market by its investment. Specifically, the base player has a stronger incentive to invest in the closed-loop case than in the open-loop model.

Theorem 5. *Suppose that there exists a closed-loop equilibrium. Then, the base capacity in the closed-loop equilibrium is at least as great as the base capacity in the open-loop equilibrium.*

Proof. Suppose that $x_p^o < x_p^c$. This relation, together with $x_p^o + x_b^o < x_p^c + x_b^c$ shown in Theorem 4, implies $2x_p^o + x_b^o < 2x_p^c + x_b^c$. Using the corollary of Lemma 9, we write

$$K_p = \sum_{s'} \lambda_p^{s'c} + \sum_{s''} \lambda_p^{s''c} \quad \text{with } \lambda_p^{s'c} > 0 \text{ and } \lambda_p^{s''c} = 0.$$

Because $\lambda_p^{s'c} > 0$ implies $\lambda_b^{s'c} > 0$, by Lemma 9 we have

$$2y_p^{s'o} + y_b^{s'o} \leqslant 2x_p^o + x_b^o < 2x_p^c + x_b^c = 2y_p^{s'c} + y_b^{s'c},$$

which proves that $\lambda_p^{s'o}$ must be greater than $\lambda_p^{s'c}$. This implies

$$K_p = \sum_{s'} \lambda_p^{s'c} + \sum_{s''} \lambda_p^{s''c} < \sum_{s}^{\prime} \lambda_p^{s'o} + \sum_{s''} \lambda^{s''} = K_p,$$

and hence, a contradiction. □

The overall outcome of this added investment is a reduction of prices compared to those prevailing in the open-loop equilibrium.

Theorem 6. *Suppose that there exists a closed-loop equilibrium. Then, the total production in the closed-loop equilibrium is larger than in the open-loop equilibrium for each time segment. Hence, prices are lower in each time segment of the closed-loop equilibrium.*

Proof. Using Theorems 4 and 5, we know that $x_p^o + x_b^o < x_p^c + x_b^c$ and $x_b^o \leqslant x_b^c$. Suppose first that $\lambda_p^{sc} > 0$. Then, $\lambda_b^{sc} > 0$ and $y_p^{so} + y_b^{so} \leqslant x_p^o + x_b^o < x_p^c + x_b^c = y_p^{sc} + y_b^{sc}$, and the result is proven for these load segments. Suppose now that $\lambda_p^{sc} = 0$ and $\lambda_b^{sc} > 0$; the two following relations hold at (y_p^{sc}, x_b^c):

$$-\alpha^s + 2\beta y_p^s + \beta x_b + \nu_p = 0, \tag{22}$$

$$-\alpha^s + \beta y_p^s + 2\beta x_b + \nu_b + \lambda_b^s = 0. \tag{23}$$

Consider a decrease of x_b from the value x_b^c towards x_b^o. Using (22), one sees that $y_p^s + x_b$ decreases

$$\frac{d(y_p^s + x_b)}{dx_b} = \frac{1}{2}$$

as well as $y_p^s + 2x_b((d/dx_b)(y_p^s + 2x_b) = 3/2)$. Relation (23) thus continues to hold with an increased λ_b^s for any decrease of x_b. Relation (22) also continues to hold until x_b hits x_b^o or y_p^s hits x_p^o. In the first case, (y_p^s, x_b^o) satisfying (22) is the open-loop second-stage equilibrium in time segment s; it satisfies $y_p^{so} + x_b^o < y_p^{sc} + x_b^c$, which proves the result. In the second case, we continue decreasing the value of x_b until it hits x_b^o, while keeping y_p bounded at its upper limit x_p^o. This results in a further decrease of $y_p^s + x_b$ until the point (x_p^o, x_b^o). This point is the open-loop second-stage equilibrium in time segments; it satisfies $x_p^o + x_b^o < y_p^{sc} + x_b^c$, which proves the result. □

The following result is a priori surprising: Even though the solutions of the open- and closed-loop equilibria may be different (and are also different from the perfect-competition equilibrium), the marginal value of the peak capacity in all time segments is the same in all these equilibria. Although the results may look strange, the underlying reason is simple: The investment criterion $K_p = \sum \lambda_p^s$ is the same in the three equilibria, and it can easily be shown that this implies the equality of the λ_p^s.

Lemma 10. *Let m, c, and o, respectively, indicate the competitive, closed-loop, and open-loop equilibria. Then, $\lambda_p^{sm} = \lambda_p^{sc} = \lambda_p^{so} = \lambda_p^s \ \forall s$ if one invests in the peak plant in the three equilibria. One has $\lambda_p^{sm} \geqslant \lambda_p^{sc} \ \forall s$ if $x_p^c = 0$ at equilibrium.*

The following theorem concludes the comparison among the different equilibria.

Theorem 7. *The total capacity and production in the closed-loop equilibrium falls between the open-loop equilibrium and the competitive equilibrium.*

Proof. Suppose first that $x_p^c > 0$. (By assumption $x_p^m > 0$.) Then, by Corollary 3,

$$K_p = \sum_{s \in S_p} (\alpha^s - \beta x_p^m - \beta x_b^m - \nu_p)$$
$$= \sum_{s \in S_p} (\alpha^s - 2\beta x_p^c - \beta x_b^c - \nu_p),$$

where $S_p = \{s \mid \lambda_p^s > 0\}$. Thus,

$$0 = |S_p| \{\beta x_p^c + \beta[(x_p^c + x_b^c) - (x_p^m + x_b^m)]\}$$

or

$$0 = x_p^c + (x_p^c + x_b^c) - (x_p^m + x_b^m),$$

and because $x_p^c > 0$,

$$x_p^m + x_b^m > x_p^c + x_b^c.$$

Suppose now that $x_p^c = 0$. Then, by Corollary 3,

$$K_p = \sum_{s \in S_p} (\alpha^s - \beta x_p^m - \beta x_b^m - \nu_p)$$
$$\geqslant \sum_{s \in S_p} (\alpha^s - 2\beta x_p^c - \beta x_b^c - \nu_p),$$

where $S_p = \{s \mid \lambda_p^{sm} > 0\}$, or

$$0 \geqslant |S_p| \{\beta[(x_p^c + x_b^c) - (x_p^m + x_b^m)]\} \quad (\text{because } x_p^c = 0)$$

or again

$$x_p^m + x_b^m \geqslant x_p^c + x_b^c.$$

By Theorem 4 we know that $x_p^c + x_b^c \geqslant x_p^o + x_b^o$, and the capacity result holds. □

By Theorem 6 we see that the production in each time segment in the closed-loop game is greater than the production in the open-loop game. To see that the competitive equilibrium has higher production than the closed-loop equilibrium, we need only consider the cases where the capacity of player p is not binding in the competitive case. First, assume that $y_p^{sc} > 0$. Then, in equilibrium,

$$-\alpha^s + \beta y_p^{sm} + \beta y_b^{sm} + \nu_p = \omega_p^{sm} \geqslant 0,$$
$$-\alpha^s + 2\beta y_p^{sc} + \beta y_b^{sc} + \nu_p + \lambda_p^{sc} = 0, \quad \text{or}$$
$$y_p^{sm} + y_b^{sm} \geqslant 2y_p^{sc} + y_b^{sc} > y_p^{sc} + y_b^{sc}.$$

For $y_p^{sc} = 0$, note that y_b^{sm} must also be zero (the marginal revenues are the same in both cases at 0 and below marginal cost) and we need only consider the case where $y_b^{sm} < x_b^m$, which leads to

$$-\alpha^s + \beta y_b^{sm} + \nu_b = 0,$$
$$-\alpha^s + 2\beta y_b^{sc} + \nu_b + \lambda_b^{sc} = 0, \quad \text{or}$$
$$y_b^{sm} \geqslant 2y_b^{sc} + \lambda_b^{sc} > y_b^{sc}.$$

Thus, the production levels are highest in competitive markets.

7. Existence and Uniqueness of the Solution of the Closed-Loop Cournot Model

Existence and uniqueness of the solution of the open-loop model were rather straightforward to establish. In the closed-loop game, these questions are much more involved and the outcomes more uncertain. We begin this section with an example that highlights the anomalous behaviors of the reaction functions in the capacity game. We explain the observed behaviors in the context of the example. Later in this section, we provide a more precise description of the discontinuities.

The numbers used in this example are roughly based on real costs from Stoft (2002). We decompose the load duration curve into nine segments of equal length (973.33 hours). We assume that $\beta = 1$ and measure the energy demand in 973.33 $*$ 10 kwh. For $s = 1$, $\alpha^1 = \$1{,}900/973.33 * 10$ kwh, and $\alpha^s = \alpha^1 - 100 * (s-1)$. The operating costs of the base and peak plants are, respectively, \$100 and \$300 per 973.3 $*$ 10 kwh. The capital costs are \$1,200 and \$400 per 10 kw of installed capacity.

To construct the reaction curve for player b, we first formulate the optimization problem for player b given x_p. This problem has equilibrium constraints. Thus, it is not a standard optimization. The model is

$$\begin{aligned}\max_{x_b}\ & \textstyle\sum_s (\alpha^s - y_b^s - y_p^s - \nu_b) y_b^s - K_b x_b\\ \text{s.t.}\ & x_b \geqslant y_b \geqslant 0, \quad \alpha^s - 2y_b^s - y_p^s - \nu_b \geqslant 0 \quad (\lambda_b^s),\\ & \alpha^s - y_b^s - 2y_p^s - \nu_p = \lambda_p^s \geqslant 0,\\ & \lambda_p^s (x_p - y_p^s) = 0.\end{aligned} \tag{24}$$

The last constraint, which is the complementarity condition for the equilibrium, removes this problem from the class of standard optimization models. Given the structure of (24), we can solve a set of optimizations to find the solution to this problem. We do this as follows.

For a given x_p, we solve (24) without the complementary condition as an optimization problem by choosing an s_p and setting $y_p^s = x_p$ for $s \leqslant s_p$ and $\lambda_p^s = 0$ for $s > s_p$. If $\lambda_p^s \geqslant 0$ and $y_p^s \leqslant x_p$ for all s, the complementarity condition is satisfied. If any $\lambda_p^s < 0$, decrease s_p, if any capacity constraint is violated, $y_p^s > x_p$, then increase s_p. Note that if the starting value of s_p is decreased (increased), then s_p can only decrease (increase) until the equilibrium constraint is met. When neither of these conditions is violated, the optimization satisfies the equilibrium constraints. To find the various points on the reaction curve where changes take place, we use line searches on x_p. The implementation is in Excel and we use Solver.

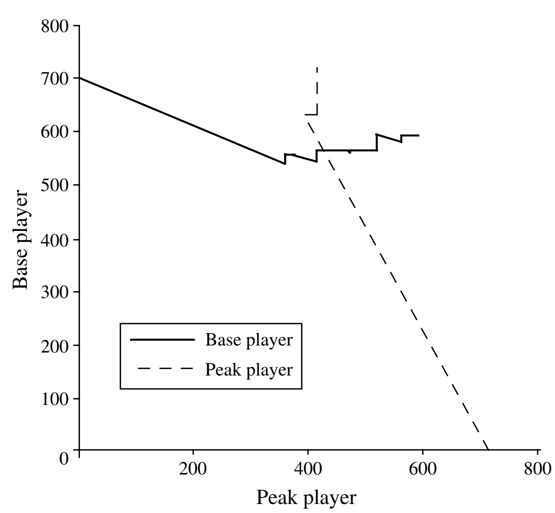
7.1. Base Player Reaction Function

Figure 4 contains both reaction functions. Before examining the causes for the shapes of the reaction functions, note that in this example, the equilibrium exists and is unique.

Clearly, neither reaction function looks like the textbook version of a reaction function. They have jumps and flat spots as well as the usual downward-sloping segments of a traditional reaction function. Also, they never reach 0, no matter the level of the other player's capacity. All of the shape anomalies make sense in the context of a closed-loop game.

The base-player reaction function is the more interesting curve. We begin with an examination of the leftmost downward-sloping portion of the curve, which has the shape of a normal reaction function. Ironically, the equilibrium cannot occur in the leftmost portion of this segment of the curve. The reason is that from our results, at equilibrium

Figure 4. Reaction curves and equilibrium for the example.



at least as many load segments have the capacity constraint binding for the base-load player as the peak player, $s_b \geqslant s_p$. In this segment of the reaction function, the levels of x_p are so low that $s_p > s_b$. By the properties of the second-stage game we developed in the previous section, in this range the roles of p and b are reversed and there is no time segment where the base-load player can force a decrease in the production of the peak player by increasing its capacity. Thus, we see the standard shape of a reaction curve in the relevant range. If the equilibrium were to occur toward the right of this segment, $s_b = s_p$ and the closed- and open-loop games are identical.

Before dealing with the complicated structures of the time segments for the middle values of x_p, we examine why the reaction curve levels out at a positive value for x_b for large x_p. Once x_p becomes large, the peak generation by player p is less than the capacity. That is, $y_p^s < x_p$ for all time segments, including the peak period. Thus, increasing x_p has no impact on y_p^s, and thereby no longer affects y_b^s for any s. This means the optimal x_b no longer changes in response to an increase in x_p and the reaction function flattens out and never reaches 0.

The discontinuous jumps in x_b happen at points x_p, where for the optimal x_b, s_p decreases to $s_p - 1$ and $s_b > s_p - 1$. At these points the marginal revenue function for player b shifts upward. To see why this effect occurs, let x'_p be the point where s_p drops to $s_p - 1$ for the x_b in the reaction curve. For $x'_p - \varepsilon$, the marginal revenue in load segment s_p for player b is

$$\alpha^{s_p} - 2x_b - x_p \tag{25}$$

because $y_p^{s_p} = x_p$.

For the x_b in (25), for $x'_p + \varepsilon$, $y_p^{s_p} < x_p$ and decreases as $y_b^{s_p} = x_b$ increases. The marginal revenue is

$$\alpha^{s_p} - (3/2)x_b - y_p^{s_p}. \tag{26}$$

Thus, at this transition point, there is a discrete increase in the marginal revenue function once $y_p^{s_p}$ begins to decrease in response to an increase in x_p. Consequently, the maximum profit occurs at a higher value for x_b with $x'_p + \varepsilon$ than for $x'_p - \varepsilon$.

Because of the changes in s_p, the profit function is piecewise concave, with the concave segments being quadratic functions. The boundaries of these piecewise concave segments are the points where s_p changes. We explore these discontinuities in greater detail later in this section.

Note that the effect of the accumulation of jumps in x_b as x_p increases is that there is no clear tendency for the reaction curve to decrease as x_p increases. A partial explanation is that as x_p increases, more time segments have $y_p^s < x_p$. When $s_b - s_p$ increases, there are more time segments in which the marginal revenue curve looks like (26), and at the transition points there are jumps that compensate for any previous decreases in x_b in the downward-sloping portions of the curve.

We have explained the extremes of the reaction function and the jumps. The two remaining features that appear in the reaction curve are the downward-sloping portions and the extended flat spots in the middle. We now show that the downward-sloping portions occur when for player b the maximum profit is achieved at a point where the profit function is differentiable with respect to x_b and x_p. Because we are dealing with the base-player optimization and not the capacity-game equilibrium, we can work with the base-player optimality conditions without imposing the equilibrium conditions for the capacity game. Taking the derivative of the marginal profit function with respect to x_b and setting it equal to zero,

$$0 = K_b + \sum_{s=1}^{s_b}\bigl[-\alpha^s + 2\beta y_b^s(x) + \beta y_p^s(x) + \nu_b\bigr] - (1/2)\sum_{s=s_p+1}^{s_b} \beta y_p^s(x). \tag{27}$$

Because we are concerned with points x_p to the right of the first jump in x_b, $s_p < s_b$ and (27) becomes

$$0 = K_b + \sum_{s=s_p+1}^{s_b}\bigl[-\alpha^s + 2\beta y_b^s(x) + (1/2)\beta y_p^s(x) + \nu_b\bigr] + \sum_{s=1}^{s_p}\bigl[-\alpha^s + 2\beta y_b^s(x) + \beta y_p^s(x) + \nu_b\bigr]. \tag{28}$$

Because $y_p^s(x) = x_p$ for $s \leqslant s_p$, (28) becomes

$$0 = K_b + \sum_{s=1}^{s=sp}\bigl[-\alpha^s + 2\beta x_b + \beta x_p + \nu_b\bigr] + \sum_{s=s_p}^{sb}\bigl[-\alpha^s + 2\beta x_b + \tfrac{1}{2}\beta y_p^s(x) + \nu_b\bigr]. \tag{29}$$

Taking the derivative with respect to x_p, we get

$$\frac{\partial x_b}{\partial x_p} = -\frac{s_p}{2s_b},$$

which is the slope of the reaction curve.

Because the profit function on the downward-sloping portion of the reaction curve is differentiable at x_b, at the flat spots the profit function for player b cannot be differentiable. The only way for this to happen is for s_b to decrease for an increase in x_b. We now explore this situation. The spot equilibrium condition for the base player when the capacity constraints do not bind on either player is

$$\alpha^s - 2\beta y_b^s(x) - \beta y_p^s(x) - \nu_b = 0. \tag{30}$$

That is, in the spot game the marginal profit from a capacity increase in this load segment is 0 for the base player at this equilibrium. In the capacity game the marginal profit from segments $s_b \geqslant s > s_p$ is as follows:

$$\alpha^s - (3/2)\beta y_b^s(x) - \beta y_p^s(x) - \nu_b > 0. \tag{31}$$

Let $y_{bu}^{s_b}$ be the solution for the unconstrained equilibrium in load segment s_b. For $x_b = y_{bu}^{s_b} - \varepsilon$, the marginal profit in this load segment is (31). However, for $x_b = y_{bu}^{s_b} + \varepsilon$, the marginal profit in this load segment comes from (30) and is 0. That is, the objective function is nondifferentiable at this point. Note, however, that the objective function for player b is concave around this point because the derivatives that exist in the neighborhood are monotonically decreasing, and at the point of nondifferentiability the right derivative is less than the left derivative.

When neither s_b nor s_p change as x_p increases, the flat spots start to slope down when x_p reaches the point where (31) is 0 in the left derivative. In our example, this happens when $x_p = 350$ and 428. Soon after x_p increases from 428, s_p decreases, returning x_b to a flat segment.

7.2. The Reaction Function for the Peak Player

In the example, the reaction function for the peak player slopes downward continuously until there is a jump followed by a flat spot. The jump and flat spot can be explained using the same reasoning as in the base-player reaction function. The flat spot happens because after a certain level, x_b is no longer binding on y_b^s for any s and the spot equilibrium no longer changes with an increase in x_b. The jump occurs because $s_b < s_p$. As this is not possible in the equilibrium, this segment is irrelevant to any meaningful analysis of the capacity game.

In the relevant range for an equilibrium the reaction function is downward sloping and behaves like a traditional reaction function because the first three load segments are the only load segments binding from the beginning to the end. Thus, there are no transition points related to changes in s_p.

7.3. Theoretical Properties of the Capacity Game at the Discontinuities

For a more detailed analysis of the discontinuities, we first introduce some notation. For a given $x = (x_p, x_b)$, use the monotonicity properties of $\lambda(\alpha)$ stated in Lemmas 3 and 4 to define

$$\begin{aligned} S_1 &= \{1, \ldots, s_1\} = \{s \mid \lambda_p^s > 0\}, \\ S_2 &= \{s_1+1, \ldots, s_1+s_2\} = \{s \mid \lambda_p^s = 0, \lambda_b^s > 0, y_p^s > 0\}, \\ S_3 &= \{s_1+s_2+1, \ldots, s_1+s_2+s_3\} = \{s \mid y_p^s = 0, \lambda_b^s > 0\}, \\ S_4 &= \{s_1+s_2+s_3+1, \ldots, S\} = \{s \mid \lambda_p^s = 0, \lambda_b^s = 0\}. \end{aligned} \tag{32}$$

We also write S_i and s_i as $S_i(x)$ and $s_i(x)$ when dependence of these elements on x is emphasized. Finally, we write

$$\Sigma = \{S_1, S_2, S_3\} \quad \text{and} \quad \sigma = \{s_1, s_2, s_3\}.$$

As a preliminary goal, we study the first-stage reaction of player b (investment x_b) as a function of the first-stage action of player p (investment x_p).

Rewriting the objective function of player b using the above sets for a given x_p, we can state that player b minimizes

$$\begin{aligned} OC_b(x_b \mid x_p) = K_b x_b &+ \sum_{s \in S_1(x)} (-\alpha^s + \beta x_p + \beta x_b + \nu_b) x_b \\ &+ \sum_{s \in S_2(x)} (-\alpha^s + \beta y_p^s(x) + \beta x_b + \nu_b) x_b \\ &+ \sum_{s \in S_3(x)} (-\alpha^s + \beta x_b + \nu_b) x_b \\ &+ \sum_{s \in S_4(x)} (-\alpha^s + \beta y_b^s + \nu_b) y_b. \end{aligned} \tag{33}$$

Define the derivative of $OC_b(x_b \mid x_p)$ with respect to x_b, where it exists as

$$\begin{aligned} MC_b(x_b \mid x_p) = K_b &+ \sum_{s \in S_1(x)} (-\alpha^s + \beta x_p + 2\beta x_b + \nu_b) \\ &+ \sum_{s \in S_2(x)} (-\alpha^s + \beta y_p^s(x) + 2\beta x_b + \nu_b) \\ &+ \sum_{s \in S_2(x)} \beta x_b B_b y_p^s(x) \\ &+ \sum_{s \in S_3(x)} (-\alpha^s + 2\beta x_b + \nu_b). \end{aligned} \tag{34}$$

It is useful to refer to $OC_b(x_b \mid x_p; S_1)$, $OC_b(x_b \mid x_p; s_1)$, $MC_b(x_b \mid x_p; S_1)$, and $MC_b(x_b \mid x_p; s_1)$, where only the first element S_1 (or s_1) is defined exogenously, independent of x, but the other S (or s) depend on x. By indexing over S_1 (or s_1), we can distinguish the convex regions of the objective function and reduce the nonconvex optimization into a sequence of convex optimizations.

The following lemma states that the objective function of player b is not convex. However, it has partial convexity properties.

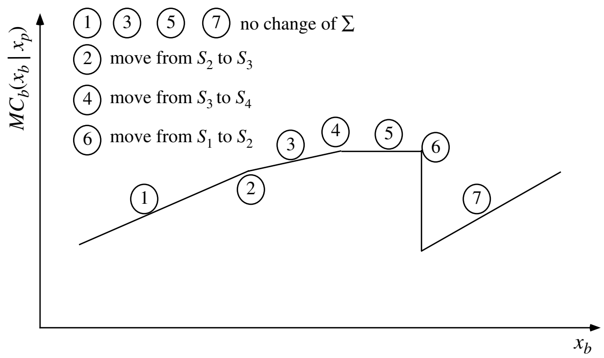
Lemma 11. *$OC_b(x_b \mid x_p)$ is a piecewise convex function of x_b for given x_p. Separation between convexity intervals occurs at points*

$$b_s(x_p) = \frac{\alpha^s - \nu_p - 2\beta x_p}{\beta}, \quad s = 1, \ldots, S.$$

The $b_s(x_p)$ identify levels of x_b where the marginal value of peak plants becomes zero. The lemma states that $OC_b(x_b \mid x_p)$ is convex in x_b as long as the sets of time segments with zero and nonzero marginal values of peak plants do not change.

The function $OC_b(x_b \mid x_p)$ is piecewise quadratic. Separation between quadratic pieces occurs when some of the $S_i(x)$ change. These changes may also create nondifferentiable points in the function $OC_b(x_b \mid x_p)$. Nonconvexities can occur only at these points. Consider the points (x_p, x_b), where this nondifferentiability of $OC_b(x_b \mid x_p)$ can occur. $MC_2(x_b \mid x_b; \Sigma)$ is still defined if one specifies the values of s_1, s_2, and s_3. Using these definitions and the proof of Lemma 11, one can state the following corollary.

Figure 5. Pattern of $MC_b(x_b \mid x_p)$.



Corollary 4. *Let x_p, x_b be a point where $OC_b(x_b \mid x_p)$ is not differentiable. Define $\sigma = (s_1, s_2, s_3) = \lim_{\varepsilon>0,\, \varepsilon\to 0} \sigma(x_p, x_b - \varepsilon)$. Then,*

$$MC_b(x_b \mid x_p; s_1 - 1, s_2 + 1, s_3) < MC_b(x_b \mid x_p; s_1, s_2, s_3\},$$
$$MC_b(x_b \mid x_p; s_1, s_2 - 1, s_3 + 1) = MC_b(x_b \mid x_p; s_1, s_2, s_3\},$$
$$MC_b(x_b \mid x_p; s_1, s_2, s_3 - 1) = MC_b(x_b \mid x_p; s_1, s_2, s_3\}.$$

Now, consider for x_p given, the evolution of $MC_b(x_b \mid x_p)$ as x_b increases. Elements of $S_1(x)$ can move into $S_3(x)$ and, similarly, elements of $S_2(x)$ and $S_3(x)$ can move into $S_3(x)$ and $S_4(x)$, respectively. Using Corollary 4, we obtain a graph of $MC_b(x_b \mid x_p)$, as depicted in Figure 5.

7.4. Discussion

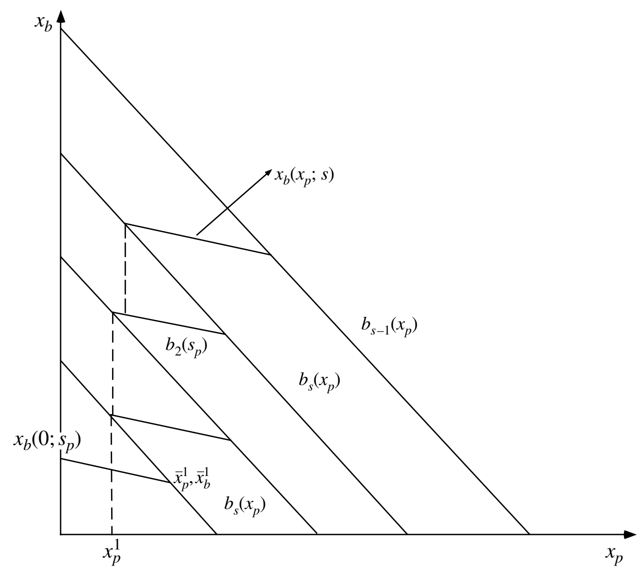
From the preceding discussion, the constraint $x_b \leqslant b_{s_1}(x_p)$ determines the set $S_1 = \{1, \ldots, s_1\}$ and defines the region of convexity of the function $OC_b(x_b \mid x_p)$. Even though $\min_{x_b} OC_b(x_b \mid x_p)$ is not a convex problem, it is piecewise convex, and $\min_{x_b \leqslant b_{s_1}(x_p)} OC_b(x_b \mid x_p) = \min OC_b(x_b \mid x_p; s_1)$ is a convex problem. The problem $\min_{x_b} OC_b(x_b \mid x_p; s_1)$ has an economic interpretation: It represents the behavior of player b when this player optimizes its capacity in the domain of x_b that makes $\lambda_p^s = 0$ (the marginal value of the plant of player p is zero) in all market segments $s > s_1$. Comparing the expressions of the objective functions, one can easily see that $OC_b(x_b \mid x_p; s_1) \geqslant OC_b(x_b \mid x_p; s_1 - 1)$. Also, $x_b \leqslant b_{s_1}(x_p)$ is contained in $x_b \leqslant b_{s_1-1}(x_p)$. Reassembling these different remarks, one can conclude that $\min_{x_b} OC_b(x_b \mid x_b) = \min_{s_1} \min_{x_b} OC_b(x_b \mid x_p; s_1)$.

While the objective function of player b is piecewise convex, as stated in the following lemma, the optimum never lies at the boundary between two zones of convexity.

Lemma 12. *The reaction of base player b (investment x_b) to the action of player p (investment x_p) can never be on a boundary $x_b = b_{s_1}(x_p)$ for some s_1.*

Proof. Suppose that the solution of $\min OC_b(x_b \mid x_p; s_p)$ is $x_b = b_{s_1}(x_p)$. By the convexity of $OC_b(x_b \mid x_p; s_1)$, $MC_b(x_b \mid x_p; s_1)$ must be nonpositive at $x_b = b_{s_1}(x_p)$.

Figure 6. Interpretation.



This implies that $MC_b(x_b \mid x_p; s_1 - 1) < MC_b(x_b \mid x_p; s_1) \leqslant 0$, and hence that player b does not select $b_{s_1}(x_p)$ as its reaction to x_p. □

The above discussion leads to a first characterization of the reaction function of player b to the investments of player p.

Proposition. *The solution to*

$$\min_{x_b} OC_b(x_b \mid x_p; s_p) \Big(= \min_{x_b \leqslant b_{s_1}(x_p)} OC_b(x_b \mid x_p) \Big)$$

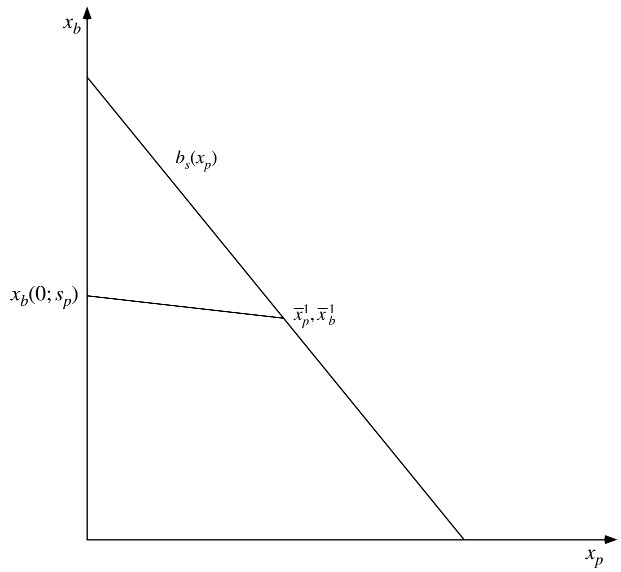
exists and is unique. Let $x_b(x_p; s_1)$ be this solution. Then, $x_b(x_p; s_1)$ is piecewise affine and continuous, with the slope of each affine segment in the interval $[0, 1)$.

The proposition suggests, but does not prove, that the overall reaction function of player b has the form depicted in Figure 6 where a particular piecewise affine function $x_b(x_p; s_1)$ constitutes the reaction function in an interval strictly between two lines $b_s(x_p)$ and $b_{s-1}(x_p)$ (because of Lemma 12). Note that at the discontinuities there are two solutions and the reaction function is not uniquely valued. To further elaborate on this intuition, consider the first segment of this reaction function.

7.5. Construction of the First Segment of the Reaction Function

Consider the initial condition, $x_p = 0$ (no peak capacity) in Figure 7. Player b reacts to this situation by solving $\min_{x_b} OC_b(x_b \mid 0)$, which is a convex problem. This solution defines the set of time segments $s = 1, \ldots, s_1^0$, where the marginal value of an investment in the peaker is positive. To define s_2 and s_3 and avoid the degeneracy $y_p^s = x_p = 0$ with $\lambda_p^s > 0$, take a perturbation $\varepsilon > 0$ of the zero capacity of equipment p. Let s_2^0, s_3^0 be obtained accordingly.

Figure 7. Construction of the reaction curve when x_p departs from 0.



Let $x_b(0; s_1^0)$ (or $x_b(0; s_1^0, s_2^0, s_3^0)$) be this solution. By construction, $x_b(0; s_1^0) \leqslant b_{s_1^0}(0)$. Consider the function $x_b(x_p; s_1^0)$, that is, where s_1^0 is kept fixed, but s_2 and s_3 are functions of point x. By the proposition, $x_b(x_p; s_1^0)$ is continuous piecewise affine with slope between 0 and -1 for each affine segment. It thus has an intersection with $x_b = b_{s_1^0}(x_p)$ because $b_{s_1^0}(x_p)$ has a slope of -2. Let $\bar{x}_p^1, \bar{x}_b^1$ be this intersection. We know that $\bar{x}_p^1$, $\bar{x}_b^1$ cannot be on the reaction function. Thus, this leads to Lemma 13.

Lemma 13. *There exists a point x_p^1 strictly between* 0 *and $\bar{x}_p^1$, where $x_b(x_p; s_1^0)$ ceases to be the optimal response when $x_p > x_p^1$. From that point on, and on some interval, the optimal response is a function $x_b(x_p; s_1^1)$ with $s_1^1 < s_1^0$. Moreover, one has $x_b(x_p^1; s_1^1) > x_b(x_p^1, s_1^0)$.*

This leads one to extend the reaction function as depicted in Figure 8. $x_b(x_p; s_1^0)$ is thus the reaction function until some point x_p^1, where s_1^0 decreases and $x_b(x_p)$ jumps by a positive amount. Let x_p^1, x_b^1 (where $x_b^1 = x_b(x_p^1; s_1^1)$) be the point after the jump; $k = 1$ denotes the first jump. This construction can be generalized.

Lemma 14. *Let (x_p^k, x_b^k) and $x_b(x_p; s_1^k)$ be the point and the reaction function obtained after jump k. One has*

$$x_b(x_p^k; s_1^k) < b_{s_1^k}(x_p^k).$$

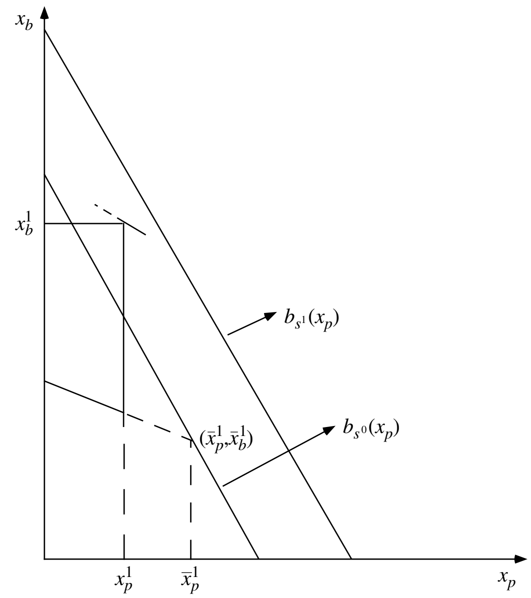
If $s_1^k > 1$, $x_b(x_p; s_1^k)$ defines the reaction function until a point x_p^{k+1}, x_b^{k+1}. At that new point, the optimal response is a function $x_b(x_p^{k+1}; s_1^{k+1})$ with $s_1^{k+1} < s_1^k$. Moreover, one has

$$x_b(x_p^{k+1}; s_1^{k+1}) > x_b(x_p^k; s_1^k).$$

This construction can be summarized in the following theorem.

Theorem 8. *The capacity reaction function of the base player in the closed-loop game is piecewise continuous with*

Figure 8. First and second segments of the reaction functions.



upward jumps. In each interval of continuity, it is monotonically nondecreasing with slope between 0 *and* -1.

Because player p sees player b at capacity whenever p is at capacity in the relevant range of its reaction function, its reaction function is continuous and monotonically decreasing with slope between 0 and -1 in the range where an equilibrium can occur. Combining the properties of the reaction functions, if they intersect, they intersect at only one point. Summing up, we obtain the following existence and uniqueness result.

7.6. Existence and Uniqueness of the Capacity Game

Theorem 9. *The closed-loop game does not necessarily have a pure-strategy equilibrium. If it does, the equilibrium cannot occur when $\lambda_p^s = 0$ and $x_p = y_p^s$. If there is an equilibrium, it is unique.*

The discontinuities have a flavor of strategic substitute and complement effects as discussed in Bulow et al. (1985). The downward-sloping affine segments reflect substitute effects. They are driven by the linear demand curves as in Dixit's (1980) model. The upward jumps look like extreme cases of complement effects where an increase of capacity of one player (here the peak) induces a simultaneous increase of the other player. It results from a reoptimization of the generation of the peak player at certain levels of peak and base capacity. This rearrangement is rooted in the discrete decomposition of the demand curve into different time segments, something that is not directly interpretable in the framework of Bulow et al. (1985).

Lemma 15. *If for $s_b = s_1 + s_2$, $x_b = b_{s_1}(x_p)$ for some s_1,*

$$\alpha^{s_b} - 2x_b - x_p - \nu_b > 0, \tag{35}$$

the reaction of the base player b, x_b, to the investment of player p, x_p, can never be on a boundary x_b for some s_1.

The reason for assumption (35) is to eliminate the possibility of a kink that creates a flat spot occurring at this point. This is equivalent to a degeneracy in a linear program. If this happens and $x_b = b_{s_1}(x_p)$ is the solution, then the kink has convexified the profit functions, and the reaction function is horizontal as discussed earlier.

8. Conclusion

This paper analyzes three capacity expansion models in the context of a restructured electricity industry. The first model assumes a perfectly competitive market as a baseline for comparison with the other models. The second model, referred to as the open-loop Cournot model, represents a market where commitments are simultaneously made on investment and sales contracts, that is, an organization based on Power Purchase Agreements. This model has the standard Cournot properties, and it is also easy to handle numerically. The third model represents an industry organized around merchant plants. It has a capacity equilibrium problem subject to equilibrium constraints on generation. This is a true two-stage equilibrium problem with nonconvexities in the first stage and is difficult to handle numerically. The nonconvexity is not surprising. Two-stage equilibrium models are extensions of bilevel and MPEC problems that are well known to be nonconvex.

To explore the different games, the models on which we have elaborated in this paper have been simplified to the case of two agents, each specializing in a particular type of plant, namely, peak and base plants. This simplified context facilitates the analysis and makes it relatively easy to identify whether there is an equilibrium, and to characterize it when it exists. The simplification also allows us to characterize the set of possible second-stage equilibria using sensitivity analysis and derive results on an a priori, badly behaved problem. This characterization can also help reduce the enumeration required to handle the nonconvexity of the problem.

The key results are as follows. The complexity of the electricity market extends to capacity expansion even without considering the difficult spatial issues. The contract market, modeled as an open-loop game, has a unique equilibrium with market prices above marginal cost, as is typical in the Cournot framework. Having a spot market partially mitigates market power as modeled in the closed-loop game, leading to quantities and prices between the competitive and open-loop models. However, the closed-loop game may not have an equilibrium. When it does, the equilibrium is unique. Because the base player has lower operating costs, in the closed-loop game it can take advantage of its position to expand its market share. Indeed, the peak player generates less in the closed-loop game than in the open-loop game despite the overall increase in production. This argues that the higher-cost generators may want to sell long-term contracts to mitigate the market power of baseload generators. We must temper these results by noting that spot markets are riskier than long-term markets, and long-term contracts help manage risk.

Because the base player increases its production relative to the open-loop game, in the solution to its optimization, the dual on the capacity constraint is lower than the cost of capacity. This anomaly is an illustration of the impact on duality theory of having equilibrium constraints. Because of the asymmetry in costs, the duality structure of solution to the peak player's optimization is unaffected by the equilibrium constraint. We intend to explore the implications of this anomaly in future research.

We expect that some of this analysis can be extended to more general models. In principle, the search through second-stage equilibria needs to be done by enumerating all complementarity sets of the second-stage problem. In the one-node, two-technology case, we were able to reduce this to a search of at most S sets. However, this may be an impossible task for a general problem with several agents controlling several technologies or when agents are spatially distributed on a grid. One longer-term objective of the paper is to show that this enumeration can be reduced by sensitivity analysis. Also, we expect that economic intuition could help develop this sensitivity analysis and characterize the nature of the relevant nonconvexities. One next step in the research will include exploring which sensitivity properties can be retained in a more general context to reduce the enumeration.

Sequential games pervade all electricity-restructuring experiences even though the literature remains relatively underdeveloped. Most of the attention in the area thus far has concentrated on the contract market (e.g., Green 1999, Newbery 1998, Wolak 1999, Bessembinder and Lemmon 2002) or multisettlement systems (Kamat and Oren 2004). The subject that has garnered the most attention is the extent to which forward markets reduce market power and the incentive of players to engage in these contracts. This problem finds its academic origin in Allaz (1992) and Allaz and Vila (1993). It has been highlighted recently by the contrast between the California debacle (where these contracts were forbidden) and the good performance of the British reform (where they were allowed). We look at a somewhat complementary problem, as we do not consider the forward/spot markets, but compare two situations that differ by the existence of a spot market.

The results presented here do not include a futures market. We have preliminary results that show the effect of a futures market in the presence of capacity restraints. These results show that the futures story is more complicated with capacity constraints, and a futures market does not necessarily have the same beneficial effect of increasing supply as in the Allaz and Vila (1993) model without capacity constraints. The multistage approach taken here can also offer insight into the maintenance games that were prominent in California.

Notation

$s = 1, \ldots, S$ load segments.
$s = 1$ peak segment.
$s = S$ base segment.
p peak player.
b baseload player.
S_p last segment for which peak capacity is the lower-cost capacity.
$i = p, b$ index of the players.
K_i investment cost.
ν_i operating cost.
x_i amount of investment by player i.
$x = (x_p, x_b)$
y_i^s operating level of player i in segment s.
p^s price in segment s.
α^s intercept of the demand curve.
ω_i^s dual on the operating constraint for segment s.
λ_i^s dual on the capacity constraint for segment s.
$y_i^s(x)$ short-term equilibrium as a function of capacity.
$\omega_i^s(\alpha^s)$ dual on the operating constraint as a function of the demand-curve intercept.
$\lambda_i^s(\alpha^s)$ dual on the capacity constraint as a function of the demand-curve intercept.
$s_i(x)$ $\max\{s \mid y_i^s = x_i\}$.
$S_i(x)$ maximum segment index for which capacity of type i is binding.
$B_i y_j(x)$ rates of change of y_j with respect to x_j's.
$y_i^s(y_{-i}^s, x)$ short-run reaction curve given the capacities.
$x_i(x_{-i})$ long-run reaction curve in the open-loop game.
$y_i(x_{-i})$ short-run solution given the other player's capacity in the open-loop game.

Appendix

Appendices 1 and 2 are available in the online companion at http://or.pubs.informs.org/Pages/collect.html.

Acknowledgment

This research was partially funded by NSF grant ECS 032536.

References

Allaz, B. 1992. Oligopoly, uncertainty and strategic forward transactions. *Internat. J. Indust. Organ.* **10** 297–308.

Allaz, B., J.-L. Vila. 1993. Cournot competition, forward markets and efficiency. *J. Econom. Theory* **59** 1–16.

Bessembinder, H., M. L. Lemmon. 2002. Equilibrium pricing and optimal hedging in electricity forward markets. *J. Finance* **57**(3) 1347–1382.

Bulow, J., J. Geanakoplos, P. Klemperer. 1985. Holding idle capacity to deter entry. *Econom. J.* **95** 178–182.

Chuang, A. S., F. Wu, P. Varaiya. 2001. A game-theoretic model for generation expansion planning: Problem formulation. *IEEE Trans. Power Systems* **16**(4) 885–891.

Daxhelet, O., Y. Smeers. 2001. Variational inequality models of restructured electric systems. M. C. Ferris, O. L. Mangasarian, J.-S. Pang, eds. *Application and Algorithms of Complementarity.* Kluwer, New York.

Dixit, A. 1980. The role of investment in entry deterrence. *Econom. J.* **90** 95–106.

Dixit, A., R. S. Pindyck. 1994. *Investment Under Uncertainty.* Princeton University Press, Princeton, NJ.

Fudenberg, D., J. Tirole. 1991. *Game Theory*. MIT Press, Cambridge, MA.

Gabszewicz, J. J., S. Poddar. 1997. Demand fluctuations and capacity utilization under duopoly. *Econom. Theory* **10**(1) 131–147.

Green, R. J. 1999. The electricity contract market in England and Wales. *J. Indust. Econom.* **47**(1) 107–124.

Grenadier, S. 2000. *Games Choices: The Intersection of Real Options and Game Theory.* Risk Books, London, UK.

Harker, P. T., J. S. Pang. 1990. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Math. Programming* **48B** 161–220.

Haurie, A., F. Moresino, O. Pourtallier. 1999. Oligopolies as dynamic games: A computational economics perspective. P. Kall, H.-J. Lüthi, eds. *Operations Research Proceedings 1998*. Springer, Berlin, Germany, 41–52.

Hobbs, B. F. 2001. LCP models of Nash-Cournot competition in bilateral and POOLCO-based power markets. *IEEE Trans. Power Systems* **16**(2) 194–202.

Hunt, S., G. Shuttleworth. 1996. *Competition and Choice in Electricity*. John Wiley and Sons, New York.

Janssens de Bisthoven, O., P. Schuchewytsch, Y. Smeers. 1988. Power generation planning with uncertain demand. Y. M. Ermoliev, R. Wets, eds. *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Berlin, Germany, 465–480.

Joskow, P. 1998. Electricity sectors in transition. *Energy J.* **19**(2) 25–52.

Kamat, R., S. S. Oren. 2004. Two-settlement systems for electricity markets under network uncertainty and market power. *J. Regulatory Econom.* **25**(1) 5–37.

Kreps, D., J. Scheinkman. 1983. Quantity precommitment and Bertrand competition yield Cournot outcome. *Bell J. Econom.* **14** 326–337.

Louveaux, F. V., Y. Smeers. 1988. Optimal investments for electricity generation: A stochastic model and a test-problem. Y. M. Ermoliev, R. Wets, eds. *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Berlin, Germany, 33–64.

Luo, Z.-Q., J. S. Pang, D. Ralph. 1996. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, UK.

Massé, P., R. Gibrat. 1957. Application of linear programming to investments in the electric power industry. *Management Sci.* **3**(1) 149–166.

Murphy, F. H., A. Soyster. 1983. *Economic Behavior of Electric Utilities*. Prentice-Hall, Englewood Cliffs, NJ.

Newbery, D. M. 1984. The economics of oil. F. van der Ploeg, ed. *Mathematical Methods in Economics*. John Wiley and Sons, New York.

Newbery, D. M. 1998. Competition, contracts and entry in the electricity spot market. *RAND J. Econom.* **29**(4) 726–749.

Ronn, E. I. 2002. *Real Options and Energy Management: Using Options Methodology to Enhance Capital Budgetting Decisions*. Risk Books, London, UK.

Schmalensee, R. C. 1981. Economies of scale and barriers to entry. *J. Political Econom.* **89** 1228–1238.

Selten, R. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *Internat. J. Game Theory* **4** 25–55.

Sherali, H. A., A. Soyster, F. H. Murphy. 1982. Linear programming based analysis of marginal cost pricing in electric utility capacity expansion. *Eur. J. Oper. Res.* **11** 349–360.

Smit, H., L. Trigeorgis. 2004. *Options and Games.* Princeton University Press, Princeton, NJ.

Spence, A. M. 1977. Entry, capacity, investment and oligopolistic pricing. *Bell J.* **8** 534–544.

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

Trigeorgis, L. 1996. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. MIT Press, Cambridge, MA.

Wei, J.-Y., Y. Smeers. 1999. Spacial oligopolistic electricity models with Cournot generators and regulated transmission prices. *Oper. Res.* **47**(1) 102–112.

Wolak, F. A. 1999. An empirical analysis of the impact of hedge contracts on bidding behavior in a competitive electricity market. Working Paper w8212, National Bureau of Economic Research, Cambridge, MA.