IS INVESTMENT IN PEAK GENERATION ASSETS EFFICIENT IN A DEREGULATED ELECTRICITY SECTOR?

Benoît Peluchon
MATISSE University Paris 1, Boulevard de l’hospital 75013, Paris, France

Abstract

Electricity markets have known turbulent times the last two years, from California crisis to nordpool’s present winter. Very high prices, with no evident link to costs of production, have been recorded in exceptional demand peaks. Each time a lack of generation capacity has been put in evidence. During December 2001, this has even lead to failures in Spain. This paper seeks to explain this by focusing on peak capacity investments. Peak generators have a crucial role during peak demand times, since they prevent any failure in the delivery of electricity. Though, their financial viability is very random since they produce electricity only when demand is exceptionally high. Under investment in peak generators may thus be underlying the peaks in prices.

Thus a model derived from electricity markets models (Green and Newberry 92, Borenstein and Bushnell 99) is build, introducing randomness and the investment decision in the picture. It allows to understand why a firm may not be willing to invest as much as a public monopoly and why market power may be more difficult to measure than by computing mark up rates.

1. INTRODUCTION

From California’s hot 2000 summer to Europe’s hard 2003 summer, price spikes are now a common occurrence on electricity markets. Though high prices are a normal phenomenon during demand peaks, those spikes have been regarded mainly as an evidence of dysfunctional markets. Firstly, they seemed to bear no link with the costs of (peak) production (1000 €/MWh in France on the 11th of august). But more important, the possibility of failures in the supply of electricity rose sharply, as reserve margins were very low. Each time a price spike has been recorded, a very high demand was coupled with low remaining capacities, even negative ones when shortcuts had to be done (California 2000, Spain 2001, Italy 2003, France 2007). Still, that there was too much production capacity in Europe was the general opinion three years ago.

The non storability of electricity is crucial here. To meet a given demand, there must be at least the same level of supply. But, as electricity consumption is cyclical through the year, the higher levels of demand are supplied by generators with low fixed costs and high proportionnal costs, peak units. Under investment in peak generation units may thus be an explanation for the recent crises, since peak generators are used to balance supply and demand. They work as an insurance against shortcuts for a given electric system. The need for peak generation depends on the per-
ception producers have about the distribution of probability of demand. It is subject to changes since this distribution changes too with consumption habits (the development of air-cooled devices for example). Because they produce only when demand is very high, their profitability is very random and may be problematic. Moreover, as many economists have already observed, electricity markets tend to function badly when demand is exceptionally high (Joskow 2001). One of the reasons is that competition for peak demand is less sharp, as many producers are already using all of their capacity. Thus, producers that have some remaining capacities have more market power and are able to get higher prices. Under investing may then reward higher profits.

This paper is an attempt to characterize and explain this under investment in peak generation capacity. Investment and randomness are introduced in a model very similar to electricity market models, such as those developed by Green and Newberry (1992), Green (1996) or Borenstein and Bushnell (1999). These models have been used to describe the price mechanism on wholesale electricity markets and to investigate the possibility of market power on such markets. Their conclusions are that wholesale prices are higher than marginal costs and they have pointed situations where firms had market power (the Pool in 1992, California in 1999). But market power is considered exceptional (Wolfram 1999) or can be attenuated by adequate regulation (Bushnell, Borenstein, Knittel 1999). Including the investment process seems to point a deeper and more structural flaw in electricity markets.

2. INVESTMENT IN GENERATION CAPACITY

2.1 The problem of peak investment

Investment in the generation industry is a very complex phenomenon. This industry has been progressively “deregulated” in Europe in the 90’s, but has largely thrived on the generation capacity built by the public national monopolies that prevailed before. At the exception of England, where since the beginning of the 90’s many gas fired units have been installed, the introduction of market mechanisms has been done in a context free of investment (at the exception of renewable energy units). Still, the growth of consumption coupled with the divestures of many generators in Europe is changing the picture. The need for new investments is growing. We do not know how the frame of institution put in place to regulate the new markets are able to deal with this new context. Investment seems to be an interesting subject to investigate and has not been treated as such by the academic literature. Peak capacity investment, especially, seems quite problematic. An investment in a base generation plant is a decision that requires forecasting the base future prices. An investment in a peak generation plant requires much more information as the peak prices depend from the base prices as well as from the future investments in every kind of generation capacity. The revenue generated by a peak plant is much more hazardous than a base plant, since it produces electricity only when every other plant produces to full capacity or cannot produce. In the same way that an option said to be “out the money”, peak plant has a value that may change drastically with any change in the way demand evolves. Take for example a generator whose proportional cost puts at the higher end of the load curve, and suppose that such a demand happens with a mean of one in ten years. This means that, on mean, investing in such a unit is profitable, but will not produce nine years out of ten. Hence our interest in how firms may deal with this problem.

2.2 Which level of investment is “optimal”? 

As already told, peak capacity has a role of insurance for the whole electric system. It means that there must be enough installed peak capacity to ensure a low shortcut probability. But what
is the price of that insurance? And what would be an “optimal” level of capacity from the point
of view of the regulator? A possible answer is to define a public “cost of failure”, that is the
public cost of not supplying one kWh of energy. Traditional French economic calculus used this
cost to compute an optimal set of generators by minimizing the total cost of supplying demand.
The probability of a shortcut, which depends on the capacity installed, has then an economic
grounding. It is the probability that gives an expected cost of shortcut equal to the cost of build-
ing one more unit of capacity. We may define thus an optimal level of peak capacity. In the
model, we make the comparison between the investment policy of two regimes: a public mo-
monopoly, whose objective is to minimize its costs, and a deregulated industry, where firms maxi-
mise their profits. It allows us to get some ideas about the strategies firms may implement and
how to judge these strategies. This paper shows that generation firms tend to invest less than a
public monopoly, in order to raise their profit.

3. THE MODEL

3.1 General features

In order to test this intuition about peak investment, a model of the market for peak energy has
been built. Following Scheinkman and Kreps (1983), we suppose the investment decision is
taken before the demand is known, then the price is fixed by the equality between supply and
demand.

The model limits itself to peak demand, that is the residual demand of energy once all the pre
peak generators produce with all of their available capacity. This implies that every change in
pre peak capacity must be included as a change in residual demand, since it lowers by the same
level the length of peak demand. We suppose the time to be reduced to a single point, we thus
set aside the possibility that power may be lacking. But including that possibility in a model
suppose to modelize the whole load curve. We focus instead on the possibility of energy deficit.
Peak demand is assumed to follow a normal distribution. We can thus write peak demand at
time \( t \) as the sum of a mean and a random shock of mean zero and variance \( \sigma^2 \):

\[
D_t = m_t + \epsilon_t.
\]

There is no price in this function. We don’t assume that consumption is inelastic to price, even
if elasticity is very weak. But we make the assumption that the value of elasticity is unknown
and that it manifests itself by an exogenous price determination in the case of the duopoly, while
there is no elasticity in the case of the public monopoly on behalf of tarification. The random
part is an oversimplification of reality, but allows us to introduce randomness and still derive
general results.

There are three different costs: the cost of “failure” (not supplying one unit of electricity), the
cost of fuel that we admit to be completely proportional and the cost of investment (cost of
building one unit of capacity). The set of variables is:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of failure</td>
<td>( C_D )</td>
</tr>
<tr>
<td>Cost of fuel</td>
<td>( C_p )</td>
</tr>
<tr>
<td>Cost of investment</td>
<td>( 3.1.1.1.1.1 )</td>
</tr>
<tr>
<td>Capacity installed</td>
<td>( Q )</td>
</tr>
<tr>
<td>Under-capacity price</td>
<td>( P )</td>
</tr>
</tbody>
</table>
3.2 The public monopoly

The objective of the monopoly is to minimize its expected total cost which is composed of three parts: investment cost, cost of production, cost of demand that has not been supplied. Its variable of decision is the level of capacity it wishes to build in order to meet demand.

The firm’s objective can be written as:

\[
\min_{\overline{Q}} E[\text{cost}] = \min_{\overline{Q}} E\left[\overline{Q} + C_P D 1_{\{D < \overline{Q}\}} + (C_P \overline{Q} + C_D (D - \overline{Q})) 1_{\{D > \overline{Q}\}}\right]
\]

where \(1_{\{D < \overline{Q}\}}\) is a function that yields 1 if the condition \(D < \overline{Q}\) is true; and 0 otherwise.

It is not possible to get an explicit expression of the optimal level of investment, but the implicit equation that defines it permits the comparison between the two regimes. This equation is:

\[
1 - P(\varepsilon < \overline{Q} - M) = 1 - P(D < \overline{Q}) = \frac{I}{C_D - C_P}
\]

This traces the link between the different costs and the probability of under-capacity. Second order condition is verified. The normal distribution of demand allows us to compute the impact of a positive variation of exogenous variables. A rise in the cost of failure implies a positive variation of optimal capacity. A rise in the cost of fuel or the cost of investment have the opposite effect. A rise of the demand trend (M) implies a positive variation of capacity so that the probability of under-capacity remains constant.

3.3 The private duopoly

We represent the regime of a deregulated industry by a symmetric duopoly. Each firm in the duopoly maximizes its profit. The price mechanism in the duopoly is very simple. If demand is lower than the installed capacity, then competition between firms yields a Bertrand outcome, and price is equal to proportional cost. If demand is greater than the installed capacity, then each firm holds a kind a market power. If one of the firm produces with all of its capacity, then the other firm is in the situation of monopoly on the remaining demand. We suppose then that a kind of elasticity takes place, so that all excess demand is cleaned by an exogenous price greater than the proportional cost that we will note as “under-capacity price”. This last price is of course lower than the cost of failure. This mechanism is very near from Kreps and Scheinkman (1983) model where capacity constraints in a two period game with price competition yields a Cournot equilibrium. But their model does not exhibit pure strategy equilibriums in case of “over-capacity”, it can not be applied to the electricity industry. We suppose firms are risk neutral. This is not a realistic assumption, but supposing they are risk averse would only strengthen the results.

The firm’s expectation of profit can then be written as:

\[
\max_{\overline{Q}_i} E[\pi_i] = \max_{\overline{Q}_{i}} E\left[(C_P \overline{Q}_i - C_P \overline{Q}_i - P \overline{Q}_i) 1_{\{\varepsilon < \overline{Q}_i, \overline{D} > \overline{Q}_i\}} + (P \overline{Q}_i - C_P \overline{Q}_i - I \overline{Q}_i) 1_{\{\varepsilon > \overline{Q}_i, \overline{D} > \overline{Q}_i\}}\right]
\]

Equilibrium is defined by the intersection of the function of reaction of each firm. The following equation gives the function of reaction of firm i:
Pi \quad CP \
\quad I 
\quad Q
\quad Q
\quad -
\quad \partial
\quad \partial
\quad -
\quad A
\quad A
\quad 1

where A is the probability of over-capacity (P(D < Q_i + Q_j))

Because equilibrium is symmetric, the equation also defines the optimal level of investment of
each firm. Equilibrium is unique and second order condition is verified on a “reasonable” range
of values: total (peak) capacity must be inferior to a value defined by the parameters of de-
mand. This value is very high and simply stresses the fact that if capacity is too great then fixed
costs can not be recovered.
The optimal capacity of a firm bears the same relationships with exogenous factors than the op-
timal capacity of the duopoly, the exogenous price assuming the role of the cost of failure.

3.4 Comparison
We are now able to compare the optimal level of investment in each regime. The two equilib-
riums are defined by:

1 - \bar{Q} \frac{\partial A}{\partial \bar{Q}_i} - A = \frac{I}{(P - C_p)} , for the duopoly

1 - A = \frac{I}{(C_D - C_p)} , for the monopoly

The two expressions are very similar, the only difference being the derivative of the probability.
It is possible to show that optimal level of capacity defined by the monopoly equation is greater
than the total capacity of the duopoly (two times the optimal capacity of a firm). This is true
even if the over-capacity price equals the cost of failure, but on the range of values on which the
duopoly equilibrium is unique.

4. RESULTS

4.1 The mechanism of under-investment
The model tells us that the level of investment of the duopoly will always be lower than the
level of the monopoly. Even if the price in case of demand in excess is equal to the cost of fail-
ure, the duopoly invests less. The fundamental difference between the duopoly and the public
monopoly is that what remains a cost for the monopoly, energy not supplied, raises prices and
profits for the duopoly. If a firm builds less capacity it lowers the quantity of electricity it may
sell, but it raises the price of this same electricity. Of course competition between firms limits
this effect, but not enough to compensate for it.

Would the result be different if there was more firms? It does not seem: if we suppose more
firms compete for peak demand, the under-capacity price may be lower than in the case of the
duopoly if total peak capacity is unchanged. But, as profitability would be lower, firms will in-
vest less, so that the probability of under-capacity rises.

What is crucial here is that peak generators have to sell power to a price higher than their mar-
ginal cost in order to finance their fixed costs. As Borenstein (2000) shows, a price equal to the
marginal cost of the system can be sufficient to contribute to fixed costs of base and half base
generators. Peak generators must benefit from market power, or they have to get some “fixed” revenues to balance fixed costs. Borenstein indicates that this is the case in California where “stand by payments” (ie to be paid for getting ready to produce in case of demand peak) exist. But those last revenues are generated by intervening on reserve markets, where unexpected variation in demand are supplied in real time. Some peak generators may not get much from those markets since they remunerate speed, not capacity in itself. And no revenues are generated for peak generators when demand is correctly expected on a day to day basis. Furthermore, the total cost of investment (building the generators and “stand by payments”) in peak capacity can be higher than simply building the capacity. Reserve markets may not be enough to tackle the problem of peak investment.

Another mechanism is explicitly designed to prevent situations of under-capacity: a (high) “crisis” price is imposed on the balancing market in order to incite producers. It may not be efficient since, as the model shows, even if the price is equal to the cost of failure (a very theoretical situation since consumers are then indifferent between consuming or not) the installed capacity by the duopoly is lower than that installed by a public monopoly. All seems to work as if there was a kind of in-built mechanism in electricity market that induce producers to limit their installed peak capacity in order to raise prices. Every disposition designed to lower market power diminishes at the same time the incitation to invest.

4.2 Market power and its measures

Those results also tell us that market power may be more difficult to measure than we think. In our model we have two distinct situations: one of under-capacity and market power, the other of over-capacity with a price equal to marginal cost. Over a long period of time, the two situations appear. A measure of market power by estimating the mark up rate over a year (as done by Wolfram (1999) in order to compare different years) can then be biased. If it is measured over a hot year, demand is lower than a “normal” year, and over-capacity prevents the existence of market power. On the opposite, a cold year will yield more tension on prices because of a lack of capacity during demand peaks, thus giving higher rates. Mark up rates are relevant if measured in “normal” demand conditions (from a statistical point of view). Of course, this is a much more difficult operation, which requires modelizing the whole load curve and its probability distribution. The conclusions of investigations similar to Wolfram (1999) and especially comparisons between rates measured on different years, are then less instructing.

Other difficulties arise when we consider that many unexpected events alter the way demand evolves. For example, the development of air-cooled devices in France have made summer electricity consumption a lot more temperature dependant. This is one of the feature that has provoked the august 2003 price spike in Europe. Such an unexpected development is only perceived when temperature effectively provokes a rise in consumption, that is at the worst moment. This explains why, though the general opinion two years ago was that too much capacities were installed in Europe, price spikes happened. If we consider that peak producers market power is relative to potential demand, then we may not be able to record it or only too late. This is a strong source of concern for regulators.

5. CONCLUSION

Peak investment is a crucial problem to ensure a good supply of energy in a country. It also seems to be a growing problem, as recent crisis on wholesale markets in Europe have shown. The model presented here is an attempt to show that deregulated systems may have difficulties to cope with it, since private firms have a strong interest in limiting peak capacities. And this problem is all the more sharp that being able to observe market power on peak supply market
requires a very good knowledge of the way consumption evolves. Computing mark up rates without that information may lead to wrong conclusions. Furthermore, mechanisms whose object is to deal with peak problems (reserve markets, crisis prices on balancing markets) does not seem adapted to deal with peak investment.

The model also stresses the limits of traditional electricity markets models. By not taking into account capacity constraints and randomness they ignore very important characteristics of the generation industry. Comparing alternative regulations on the unique basis of those models is likely to induce mistakes. Of course, the present model is very limited itself, especially concerning the pricing mechanism. It does not include dynamics, nor does it examine the effects of the threat of entry. And though it shows that a public monopoly better deals with peak investment than a private duopoly, it is not enough to conclude to the superiority of the later over deregulation. It is way too simple to integrate all the parameters we have to take into account in order to compare the two regimes. But, as an extension of traditional models it helps to put a light on the biases of theory in order to correct them.

We have a strong need to enhance our knowledge of how a deregulated electricity industry tackle the problem of investment. As recent examples show, it is not likely that it always does well. It would be a strong contribution to the public debate about industry regulation and energy industry. If we do not, and if regulators do not handle the problem, the future is prone to bring new crises.

REFERENCES


