1 Security of supply of electricity

Problem analysis and policy framework

1.1 Introduction

This chapter provides an analysis of the question of generation adequacy in competitive electricity markets.[1] Secondly, a policy framework is developed for selecting among the policy options to maintain generation adequacy. The existing generation stock is considered adequate if it can be expected to meet demand under all reasonable conditions, considering normal outage rates. Concerns whether competitive electricity markets provide a sufficiently strong and early enough investment signal rose after the crisis in California's electricity market in 2000 and 2001. Shortages in other places, such as in New Zealand, Scandinavia and, most recently, Italy, have fueled these concerns. A number of adjustments to the market design have been proposed with the purpose of stabilizing the volume of generating capacity. A systematic framework for the selection of such a *capacity mechanism* has not been developed yet, however.

Focus on Europe

The focus of this chapter is upon Europe, because European electricity markets have several specific features. First, most European markets do not have a mandatory power pool. Market parties may sell their electricity bilaterally and only need to notify the system operator of their physical programs. Second, many European markets have significant trade volumes with neighboring markets, while the connected market models often vary greatly. Third, hydropower plays a limited role in many European markets, the exceptions being Scandinavia and the Alp countries. The latter factor means that most European power markets are capacity-constrained, rather than energy-constrained.

Many capacity mechanisms proposed

Capacity mechanisms vary widely in the way they are intended to work and with respect to their implementation requirements. Some provide financial incentives to generating companies, while others control the volume of generating company. Some are designed for mandatory pools, which means

[1] This chapter is based on the dissertation of De Vries (2004).

they might need to be adjusted for implementation in Europe. In most cases, little attention has been given to the issue of trade: how to prevent the investment incentive from 'leaking' abroad, and how to make a capacity mechanism immune from regional shortages? This chapter develops a set of criteria to evaluate the different proposed capacity mechanisms, describes the advantages and disadvantages of the different capacity mechanisms and, most importantly, develops a framework for deciding which capacity mechanism to implement under which circumstances.

Electricity is an unusual 'product'…

1.2 Why a capacity mechanism?

1.2.1 <u>The narrow investment optimum</u>

Electricity markets have a different dynamic from other markets due to three characteristics:

- Electricity is a strongly time-limited product. It cannot be stored, other than in pumped-hydro facilities, in a commercially viable way. However, the electricity supply system can only function in a stable manner if supply and demand are continuously balanced.
- The supply of electricity is only partly characterized by a gradually increasing marginal cost function. When all available generation units are producing electricity, no marginal increase is possible in the short term. As a result, the marginal cost curve ends with a perfectly price-inelastic section.
- The demand for electricity also is highly inelastic. This may be caused by the fact that there is no readily available alternative for most applications of electricity. At least as important is, however, that few consumers receive the required price information in time to adjust their behavior. Moreover, electricity consumption is usually measured over long periods, so consumers have no incentive to shift consumption from peak hours to off-peak hours. As a result, few consumers adjust their electricity consumption to the current price of electricity, so that the observed price-elasticity of consumers is extremely low. There are multiple experiments aimed at increasing consumer price-elasticity, but in most electricity systems their impact still is small (Nilssen and Walther, 2001; Roberts and Formby, 2001; Sæle and Grønli, 2001).

… with as consequences:
- *price volatility*
- *chance of blackouts*

The combination of these three characteristics is the reason that most mechanisms which aid the clearing of other markets, such as a delay in the delivery of the good, consumers switching to other goods or higher prices leading to a reduction in demand, are not available in current electricity markets. This has significant consequences: wholesale electricity prices are highly volatile, and secondly, there is a chance of service interruptions.

Modified theory of spot pricing still holds…

With some modifications, the theory of spot pricing still holds, even if demand is assumed to be fully inelastic. The main consequence of insufficient demand price-elasticity is that there is a risk that the market does not clear, obviously in the case that physically there is not enough generation capacity available to meet demand. At this point, the market does not reach a price equilibrium and some load will need to be shed. The cost of installing so much capacity that the chance of load shedding would be reduced to zero would exceed the social cost of the load shedding which would be avoided.

Need for price cap at VOLL

However, if the market does not clear, it may be necessary to institute a price cap to protect consumers against overcharging (e.g. Ford, 1999; Hobbs *et al.*, 2001c; Stoft, 2002). If consumers are not involved in real-time price setting, they otherwise might find themselves paying more for electricity than their value of lost load. This price cap needs to be determined carefully, as it impacts the attractiveness of investment in generation capacity. The price cap needs to equal the average value of lost load (VOLL), because at this price consumers should, on average, be indifferent whether they receive electricity or not. Stoft (2002) shows that in a perfectly competitive market, this results in an optimal level of investment in generation capacity, with an optimal duration of power interruptions. Therefore the theory of spot pricing still is valid, even if demand is fully inelastic. Price caps can be problematic, however, because it is difficult to determine the optimal level, as the value of lost load is difficult to measure (Willis and Garrod, 1997; Ajodhia *et al.*, 2002).

Energy-only market vulnerable to distortions

Although theoretically sound, the reliance upon periodical price spikes to signal the need for peaking capacity has some significant weaknesses. To begin with, there is the risk that the price cap is set at the wrong level, resulting in over or under-investment. However, there are more fundamental issues. The first is that investment in peak generation units is quite risky, so that small distortions of the investment signal may have large consequences. The second is the argument that there is a positive externality associated with investment in peaking units, because security of supply is a public good (due to the network character of electricity supply). The third factor is the inevitable development of market power during periods of supply scarcity. These issues will be addressed in the next sections.

Types of market failure

1.2.2 Market Failure

Now we will discuss a number of factors which may disturb the narrow investment optimum. The following types of market failure can be discerned (based, in part, upon Hobbs *et al.*, 2001b):

- price restrictions,
- imperfect information e.g., regarding consumer willingness to pay or future supply and demand,
- regulatory uncertainty,
- regulatory restrictions to investment, and
- risk-averse behavior by investors.

Price restrictions

The fact that a price cap may be needed to protect consumers against overcharging in times of scarcity represents a significant risk, because the optimal level of the price cap is difficult to determine. While the theory is clear that the price cap needs to be equal to the value of lost load, there are many methods of measuring the value of lost load with widely varying outcomes (see for instance Willis and Garrod, 1997; Ajodhia *et al.*, 2002). The cost of erring is high. A price cap that is not equal to the value of lost load likely results in a sub-optimal level of investment in generation capacity.

Imperfect information

Producers lack the information needed for socially optimal investment decisions (Hobbs *et al.*, 2001b; Stoft, 2002). This increases the investment risk and therefore reduces the willingness to invest. In order to calculate the probability that peak units will operate and to calculate the expected return on investment, generating companies need to know both the stochastic distribution of the demand function (so they know the distribution of the frequency, duration and height of price spikes) and the expected development of total available capacity (Hobbs et al., 2001a). The exact characteristics of the demand function are difficult to estimate, especially in newly liberalized markets for which no long time sequences of empirical data are available. Moreover, the basic characteristics of demand change over time (for instance due to the introduction of new technologies) which reduces the validity of demand functions based upon historical data.

Uncertainty due to changing regulations in related areas…

Regulatory uncertainty increases investment risk and therefore adversely impacts the willingness to invest. Regulatory uncertainty can be considered as a negative externality associated with changes in public policy. Especially in newly liberalized markets such as most electricity markets, regulatory uncertainty can be a significant factor. Consider, for example, a few of the policy changes which currently are underway in Europe:

- On November 25, 2002, the European Council on Transport, Telecommunications and Energy reached political agreement on amendments to the electricity and gas Directives and a regulation for electricity (EC, 2002).
- The European gas market is in the middle of a liberalization process. Most notably the development of the gas transport tariff system, including charges for flexibility and imbalance penalties, is highly uncertain. This has a considerable impact on a business plan involving today's state-of-the-art gas-fuelled generators.
- Additionally, there is uncertainty about future European environmental rules, such as cooling water regulations or the specifics of the proposed CO_2 emissions trading scheme (EC, 2001).

… and due to the risk of a price cap

A second source of regulatory uncertainty, with an equally significant impact upon the willingness to invest, exists with respect to the question whether a period with high prices will give cause to the government or the regulator to implement a maximum price or, if a maximum already exists, to lower it. Volatile prices are not only a risk for investors, but also for regulators due to the public protests they give rise to. Most electricity systems start liberalization with ample capacity. In fact, the desire to reduce excessive reserve margins was a motivation for liberalization. If, after the initial excess capacity has disappeared, a period develops in which prices are many times higher than their historical levels, consumers may consider this a failure of liberalization and demand intervention. This occurred in San Diego at the beginning of the crisis in California, when even a brief period of high consumer prices proved politically unacceptable (Liedtke, 2000). The political risk of being held responsible for high electricity prices, whether these are economically efficient or not, translates into a risk for investors of political intervention. Hence price volatility itself brings about regulatory risk, at least until sufficient experience has been gained with liberalized markets that investors know whether they should expect political or regulatory intervention or not (Oren, 2000; Newbery, 2001).

Regulatory restrictions to investment

Obstacles to obtaining the necessary permits may be another cause of underinvestment. While the social benefits of a proper licensing process are not disputed here, it should be taken into account that they may incur negative side-effects. Firstly, the permitting process can be lengthy, thereby increasing the response time of generation investment to an increase in demand. Especially in a situation of incomplete information about the future development of supply and demand, this may contribute to investment risk. A second effect of increasing the lead time for the construction of new plant is that it may contribute to investment cycles. This subject will be further discussed below. A third effect of permits is that they may impose additional requirements on generators, leading to operational constraints to the response to market signals. An example is that cooling water regulations may restrict operation during periods of hot weather.

Risk aversion

The theoretical approach by Caramanis *et al.* assumes that generating companies behave in a risk-neutral manner with respect to investment. This is not necessarily the case, especially when many risks themselves are not well understood. Given the many unquantifiable risks in a liberalized electricity market, it is not unlikely that investors in generation capacity choose a risk-averse strategy with respect to generation investment (Vázquez *et al.*, 2002). If all investors do so, none of them lose market share, so the penalty is limited to a loss of sales during periods of supply shortage. However, this loss of volume is small, compared to overall production of electricity, and is likely to be more than compensated by the high prices that develop during a period of supply shortage. Therefore a collective strategy of risk-averse investment behavior is beneficial to the generation companies, as long as this does not attract newcomers to the market. Such a risk-averse investment strategy would lead to less installed capacity than would be socially optimal.

1.2.3 Investment cycles

Investment in response to prices arrives too late

A year before the California crisis started, Ford (1999) published a paper in which he used a computer model to show that investment in electricity generation facilities is inherently unstable in a system with rules such as in California. His explanation is that investment is not aimed at dampening business cycles, which it would do if the right amount of new capacity became available at the right time, but at making a profit. Because investors tend to wait until they are reasonably certain that they can make a profit, and because they tend to overreact (in part because they do not know their competitors' plans), Ford considers the interaction between the price signal which a power exchange provides and investment inherently unstable.

Investment cycle

Ford's argument is essentially that a combination of risk-aversion and an insufficiently long time horizon leads to a delay of investment. Due to the low elasticity of supply and demand, the price signal will not indicate scarcity until the capacity margin is so slim that the chance of service interruptions has become unacceptably large. The long lead time for new investment means that, once a shortage has developed, this shortage becomes worse before it is alleviated with new generation capacity. Ford's argument is reinforced by the argument from the previous section, that generation capacity is undervalued during periods of abundant supply and overvalued during periods of scarcity.

Cycles can be suppressed if investors can look ahead far enough

Visudhiphan *et al.* (2001) contend that investment cycles are not inevitable, as long as investors are able to anticipate market developments. However, as we saw above, sufficient information about future supply and demand is lacking. In their simulation, Visudhiphan *et al.* also find that backward looking investment, that is, investment based upon recent experience in the market, will lead to investment cycles. Stoft (2002) arrives at the same conclusion. He notes that the distribution of price spikes may be such that investors would need to have a time horizon of several decades to determine the real average revenues from price spikes. If they use a shorter time horizon, they are bound to overestimate or underestimate their expected revenues.