NOT TO BE QUOTED WITHOUT PERMISSION

Insufficient investment in generating capacity in energy-only electricity markets

Laurens De Vries
Delft University of Technology

Karsten Neuhoff
Cambridge University

August 30, 2003

Keywords: investment, electricity, consumer utility, long-term contracts

JEL-Classification: D11, D40, D52, L94

Contact Details: Karsten Neuhoff
Full affiliation: Faculty of Economics, University of Cambridge
Address: Sidgwick Avenue, Cambridge, England, CB3 9DE

Tel: ++ 44 1223 335290
Fax: ++ 44 1223 335299

Insufficient investment in generating capacity in energy-only electricity markets

Abstract:

In theory, competitive energy-only markets should provide optimal incentives for investment in generating capacity. In practice, long-term contracts develop insufficiently and most current market designs do not have institutions in place that allow private customers to be covered by long-term contracts. As a result, generating companies are restricted in their ability to transfer risks to consumers. In such an environment generation companies invest in less generation capacity while consumers would prefer to pay for the additional costs of more generation capacity then in a world with long-term contracts.

Lack of information provided by long-term contracts also creates a risk of an investment cycle, with investor decisions lagging the demand of the market. Furthermore short term energy markets are vulnerable to price manipulation during periods of scarcity. As a result, prices are expected to be higher and supply less reliable than in the benchmark case which allows to hedge risk and mitigate market power with the help of long-term contracting. A number of adjustments to the market structure have been proposed which may improve reliability and stabilize electricity prices, of which the creation of institutions to sign long-term contracts on behalf of final customer seems most promising.

1 Introduction

This paper addresses the question, whether liberalized electricity markets will invest sufficiently in generation capacity, so that the chance of electricity shortages and the resulting service interruptions remains near the social optimum. Power shortages in California, Norway, Sweden, Brazil, New Zealand and, most recently, Italy, cast an increasing doubt upon the ability of liberalized markets to meet the public demands for a stable electricity supply at affordable prices. The different cases vary widely, however, due to the specific technical and regulatory circumstances. Therefore, the available case material does not suggest obvious conclusions about the general ability of power markets to maintain the level of generation adequacy that the public expects. Moreover, most of the examples are in systems which depend heavily upon hydropower, which leaves the question unanswered whether this is the cause or whether systems without hydropower also risk shortages. The social cost of generating capacity shortages is so high, however, that a thorough analysis of the issue is in place.

The theory of spot pricing claims that under ideal conditions, electricity spot markets can provide efficient outcomes both in the short and in the long term (Caramanis, 1982; Caramanis et al., 1982). In its simplest form, the theory of spot pricing states that electricity markets are similar to other commodity markets, with the exception that the price varies continuously as demand changes from moment to moment because the product cannot be stored. Consequently, a new market equilibrium needs to be found for each moment in time, which is made possible with communications technology. The spot prices should result in an efficient dispatch and allocation of available resources and should also signal the need for additional generating capacity.

The belief that unregulated markets in electricity generation can produce an optimal outcome in the long term is widely shared (cf. Shuttleworth, 1997; Hirst and Hadley, 1999), although it appears to be waning today. Generally, this school of thought asserts that a shortage of investment in generating capacity is only caused by obstacles to the proper functioning of the market mechanism, such as price restrictions or construction permit requirements. The proper course of action, in this view, is to improve the investment climate by eliminating all extraneous sources of risk, such as regulatory risk, and other obstacles to investment.

There is a lack of scientific agreement about the issue, however, which is reflected in the differences between existing electricity markets. Spain and a number of South American systems try to stimulate investment in generation capacity by providing capacity payments to generation (Vázquez et al., 2002). Three systems on the East Coast of the USA (PJM, the New York Power Pool and the New England Power Pool) use a system of capacity requirements (PJM Interconnection, L.L.C., 2001; see for an introduction Besser et al., 2002). Most European systems, on the other hand, have no specific provisions to ensure adequacy of capacity. Instead, they rely on the energy market to provide sufficient revenues to incentivice investment in generation capacity. We will present arguments that cast doubt on the ability of energy-only markets to provide an adequate level of generating capacity. The main failure we identify is that final consumers can not sign long-term contracts with generation companies. This eliminates the opportunity to hedge risks. Generation companies react by reducing the equilibrium invested quantity. Consumers on the other side would prefer additional generation capacity put in place. To bridge this gap we suggest to reinstate some franchise that is responsible for signing long-term contracts. A number of additional factors are discussed which could contribute to the failure of energy only markets to provide for sufficient investment in generation capacity, e.g. investment cycles, regulatory uncertainty, market power.

This paper is structured as follows. The next section briefly recapitulates the theory of spot pricing as the starting point of the analysis. Section 3 argues that even a perfect market will not lead to an optimal equilibrium level of investment, from the perspective of consumers, in the presence of uncertainty. Moreover, risk-aversion by consumers and/or producers will worsen the situation. Section 4 shows why the current market design does not allow for long-term contracting. Section 5 presents additional sources of uncertainty that could contribute both to a reduction of equilibrium investment in generation capacity and to investment cycles (Section 6). The effects of market power are discussed in Section 7. Section 8 discusses capacity mechanisms to increase investment, and Section 8 reflects upon the impact of inter-system trade. We conclude with Section 10.

1 The theory of spot pricing

There is extensive discussion whether spot pricing will provide for sufficient investment in electricity systems to secure reliable electricity supply. Caramanis et. al. (1982) show in a theoretical model that a well-defined, unregulated market will provide for sufficient investment in generation capacity. We ask, whether the assumptions underlying the result are satisfied - particularly relating to the behavior of investors and consumers in the presence of uncertainty.

One requirement frequently quoted for the success of liberalized energy markets is that demand is sufficiently price-elastic so that the supply and demand functions always intersect. In practice, the observed price-elasticity of electricity demand is quite low for a large portion of the demand curve. The production of electricity also is characterized by a partly inelastic supply function. When all available generation units are producing electricity, no increase is possible in the short term. As a result, there is a chance that the supply and demand functions do not intersect. Physically, a mismatch between supply and demand is not allowed, as it would cause technical instability of the system. As electricity cannot be stored in a commercially viable way, other than in pumped-hydro facilities, this means that whenever demand threatens to exceed supply, the system operator needs to intervene. By interrupting electricity service to groups of customers, he artificially balances supply and demand.

When the market does not clear, it is necessary to institute a price cap. A price cap is necessary to protect consumers against overcharging (e.g. Ford, 1999; Hobbs et al., 2001b; Stoft, 2002). If consumers are not involved in real-time price setting, they otherwise may find themselves paying more than their value of lost load (VOLL). The 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 VOLL, because at this price consumers should, on average, be indifferent whether they receive electricity or not. Estimations suggest that the value of lost load is some two orders of magnitude higher than regular electricity prices. Accurate estimations are difficult because there are many methods of measuring VOLL with widely varying outcomes (Willis and Garrod, 1997; Ajodhia et al., 2002).

Stoft (2002) shows that in a perfectly competitive market, a price cap equal to the average value of lost load 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. This is the case even though rotating black-outs cause inefficient allocation: some consumers would prefer to pay a higher price to ensure uninterrupted electricity supply, while others would prefer more frequent interruptions if that would lower their electricity bill. However, even a public enterprise with benevolent management would face the same dilemma. It can only be resolved with real-time metering or active demand side management. It is unclear whether the installation and transaction costs of such technology are justified by the increased allocative efficiency.

A consequence of the low price-elasticity of demand and the finite length of the supply curve is, however, that wholesale electricity prices can be very volatile. During price spikes, when prices reach the level of the price cap, peaking units can recover their costs. If they occur often enough, price spikes will attract new investment. In a perfect market, a long-term equilibrium will develop between the occurrence of high prices and investment in new capacity. The government can influence the investment incentive by adjusting the maximum price.

2 Risk aversion without long-term contracts results in under-investment

The theoretical approach by Caramanis (1982) assumes that generating companies and consumers behave in a risk-neutral manner with respect to investment. This is not necessarily the case. We will show that if final customers can not sign long term contracts with generators to cover their peak demand, then risk aversion results in under-investment, which leads to higher prices, lower adequacy and therefore lower reliability.

In this Section, we will assume that consumers cannot sign long-term contracts for generation capacity. (In the next section we will return to this assumption.) Contracts allow the generating companies to shift the risk caused by uncertainty to consumers. In a model world with risk-neutral agents, the presence of uncertainty does not change the investment and consumption decisions; therefore the lack of long term contracts does not matter. Then a competitive and transparent market with rational investors provides for the optimal amount of generation capacity.

However, risk aversion of consumers and generators means that a market without long-term contracts will provide for less generation capacity than a market with long-term contracts. We will demonstrate this for three types of uncertainty: first, uncertainty regarding future aggregate demand e.g. due to a deviation of climate from the average like an unexpected cold winter or hot summer; second, a deviation of industrial demand from the predicted growth path; third, a change of individual demand.

2.1 The role of demand price-elasticity

One of the causes of the high volatility of electricity markets, and hence of the high investment risk, is the low price-elasticity of demand. In many markets, the demand for electricity arguably is highly inelastic, as there are few substitutes. Institutional arrangements probably further reduce demand price-elasticity by limiting incentives and opportunities for consumers to respond to real-time prices.

Vertically integrated utilities either build their own generation plants (sometimes in cooperation with neighboring utilities) or sign long-term contracts for generation capacity. Usually, the costs are passed on to consumers in the form of a fixed tariff, so that consumers do not adjust their demand in times of electricity shortage. This is not necessarily the case, however: there are examples of monopolistic utilities where the regulatory environment induced them to stimulate demand responsiveness. This can be achieved by linking the electricity tariff directly to wholesale prices, but other options exist, too. It suffices that the marginal consumption is exposed to the wholesale market price. Demand-side management programs, which allow savings if some load can be interrupted, are one way; charging for deviations from the average consumption at the wholesale price level is another.

An example of the latter is a tariff that specifies that a consumer can always receive his average annual electricity demand at the fixed retail price. If the consumer reduces demand, then he will be paid for at the difference between wholesale price minus retail tariff, while if he increases demand, he has to pay according to the wholesale price for the additional energy. An extreme version of this program was successfully implemented in Brazil to combat the energy shortage of the hydro system in 2002. Consumers could receive 80% of average consumption at usual retail rates.[1] If they exceeded this level of consumption, they were first warned and subsequently disconnected. This extreme version was probably required in an society with high inequality to ensure that not only poor people, who are more price sensitive, contribute to the solution of the energy crisis. One would expect that in countries with less inequality price based mechanisms would work equally well.

Similarly, the price-elasticity of demand may vary between liberalized markets. This does not only depend upon the regulatory structure of the market as upon the structure of consumption. In Norway, in which supply companies frequently adjust the retail electricity price to match the development of wholesale prices. The Norwegian model ensures that final customers can reduce consumption in periods of scarcity. In Norway this model is particularly useful, because the hydro dominated system can face dry periods, as in 2002/2003, which require a demand-side reaction. Given that electric heating is common in Norway, switching to wood for domestic heating can save significant amounts of electricity and can be easily induced by increased retail tariffs. In systems where electricity is not used for heating (or cooling), demand typically is much less elastic.

The degree to which consumption is covered by long-term contracts may also influence the price-

[1] The website on the BR electricity crisit is: http://www.energiabrasil.gov.br/EnergiaBrasil.htm

responsiveness of demand. In principle, it should not matter, as holders of long-term contracts would be interested in reducing demand and re-selling surplus electricity if the spot price is high, but in practice their may be significant transaction costs and institutional barriers to doing so. When subsequently referring to long-term contracts, we imply that consumers are exposed to a fixed retail tariff based on the prices paid for long-term contracts. These tariffs are complemented by mechanisms to encourage demand side management as described above. When referring to a world without long-term contracts, we assume a situation where consumers are exposed to the wholesale spot prices with their entire demand.

2.2 Weather-related uncertainty of demand

2.2.1 The perspective of consumers

First assume all consumers are exposed to the same unexpected cold winter which increases energy demand, e.g. for electric heating, lighting and pumping of water circulation. The weather condition changes the monetary value consumers derive from consuming the amount C of electricity: $M(C)$.

However, it should not be represented, as it usually is, by a shift of the monetary value function to a cold winter utility function $M_C(C)$ which represents higher monetary value for the same amount of electricity: $M_C(C)>M(C)$. It is true that in a cold winter electricity is more important, as represented by $M_C(C)>M(C)$, but the overall utility of a person is not necessarily increased by the cold winter. In contrast, consumers would say that their utility is reduced if they have to live with the same amount of electricity in a colder winter, because they can satisfy less of their non-heating related electricity demand. Therefore the unexpected cold winter is represented by a reduction of the electricity available for consumption after climate related additional electricity demand ε (to simplify subsequent calculations assume $E(\varepsilon)=0$).

$$M_\varepsilon(C)= M(C-\varepsilon) \qquad < M(C) \;\; if \;\; \varepsilon>0 \tag{1}$$

The colder than expected winter therefore reduces consumers' welfare, the warmer than expected winter increases their welfare.

Assume consumers can decide how much generation capacity K to invest in or contract through long-term contracts. Without loss of generality, we normalize variable costs of generation to be 0.

The cost of generation capacity therefore is c^*K, with c the long-run marginal cost of capacity. Assuming no uncertainty ($\varepsilon=0$) consumers' 'wealth' π after investing in K and consuming $C=K$ electric energy is:

$$\pi(K)=M(K)-c^*K \tag{2}$$

Using the first order condition with respect to K we obtain an equation defining the optimal installed capacity, which we label K_W, in a world **w**ithout uncertainty about future demand:

$$c= M'(K_W) \tag{3}$$

If risk-neutral consumers face uncertainty about their future demand due to uncertainty about the weather, then they will maximize the expected 'wealth':

$$\pi(K)=E[M(K-\varepsilon)-c^*K] \tag{4}$$

The first order condition with regard to K gives the optimal investment quantity for a risk-neutral consumer perspective, labeled K_N:

$$c= E[M'(K_N -\varepsilon)] \tag{5}$$

Usually we assume that the marginal monetary value M' of energy is convex $M'''>0$, therefore $E[M'(K_N -\varepsilon)]> M'(K_N)$. Marginal monetary value of energy is also decreasing $M''<0$, therefore (3) and (5) can only both be satisfied, if $K_W<K_N$. The optimal installed capacity K_N in an uncertain world with risk-neutral consumers exceeds the capacity K_W installed without uncertainty. Stoft (2002) finds a similar result, when he introduces uncertainty regarding the availability of generating capacity.

So far we assumed consumers are risk neutral. This implies that they are indifferent between, for instance, the option to have 100 Euros spending money available every months for their household and the alternative option to have a 50% chance to either receive 70 Euros or 130 Euros. However, it appears more appropriate to assume that consumers are risk averse. (Pålsson (1996) for instance, finds that Swedish households are highly risk-averse.) Risk-averse households prefer to minimize deviations from the average, because the additional utility of one Euro in months when they have 130 Euro spending money is lower than the additional Euro in months with only 70 Euros spending money.

In (1) we introduced a monetary value function to translate electricity consumption into equivalent money income. Adding to this function other income minus expenditure we obtain some measure of wealth of the consumer in a given period.

Consumer utility is a monotonic function of this one-dimensional measure 'wealth'. If we assume that consumers are risk neutral, then utility is a linear function of 'wealth'. In contrast, risk averse agents exhibit decreasing marginal utility of an additional unit of 'wealth' with higher 'wealth' levels. This can be represented by a utility function $U(\pi)$ weighting benefit (4) in any one demand realization. Utility is increasing with 'wealth' $U'(\pi)>0$ but at a decreasing rate $U''(\pi)<0$.[2]

$$U=E[U\{M(K-\varepsilon)-c^*K_R\}] \tag{6}$$

The FOC with respect to K gives an equation for the equilibrium capacity K_R installed by risk averse consumers:

$$c = E\left[\frac{U'\{M(K_R - \varepsilon) - c * K_R\}}{E[U'\{M(K_R - \varepsilon) - c * K_R\}]} \quad M'(K_R - \varepsilon) \right] \tag{7}$$

Equations (5) and (7) differ in the weighting factor, which causes the equilibrium invested capacity to increase $K_R>K_N$. This can be explained as follows. The denominator of the weighting factor is constant and the nominator is decreasing in ε because $M'>0$ and $U''<0$. Therefore in (7) more weight is put on states of the world with negative ε which are the states with small M'. If we assume (hypothetically) that $K_R=K_N$ then the right hand side of (7) would be smaller than c. This can only be compensated by setting $K_R>K_N$. If consumers are risk averse, then they want to reduce the downside risk they face in cold winters and will therefore contract for more energy than in a world of risk-neutral agents.

---

[2] Our representation of the utility function differs from the more general utility function which combines risk aversion and decreasing marginal utility of individual goods, by stating utility as two-dimensional function of consumed energy C and discretionary income M: $V(C,-cK)$. This representation is more general, because it depicts R^2 into R while we only use two functions depicting R into R. The more general representation coincides with our representation if for all C, cK: $V_1=U'M'$ and $V_2=U'$.

2.2.2 Investors' perspective

Now let us assess the situation from the perspective of investors in a competitive market. The market clearing price will be $p=M'(K)$. In a competitive world, new investors will enter the market until the equilibrium price p equals costs c of additional generation capacity $p=c$. Combining these two equations gives $c=M'(K)$, which is identical to (3). In a world without uncertainty the market will provide for the optimal investment quantity K_W.

In a world with uncertainty but where investors are risk-neutral, they will ensure that on average they can cover their costs $c=E[M'(K-\varepsilon)]$. The exact correspondence with (5) shows that even with uncertainty the market will provide for the appropriate amount of generation investment K_N as long as agents are risk neutral.

If investors are assumed to be risk averse, then their expected benefit from investing in one unit of generation capacity is $\pi=E[U\{M'(K-\varepsilon)-c\}]$. Free entry requires that an additional investor would not benefit from providing an extra unit of generation capacity.

$$E[U\{M'(K_I-\varepsilon)-c\}]=0 \tag{9}$$

Investors are risk averse, therefore U'>0 and U''<0 and we obtain:

$$E[U\{M'(K-\varepsilon)-c\}]< U\{ E[M'(K-\varepsilon)-c]\} \tag{10}$$

Assume risk-averse investors would invest the same amount of generation capacity as risk-neutral investors K_N, then the right hand side of (10) is zero, implying that the left hand side is negative. To increase the left hand side to satisfy (9) we have to decrease K_I below K_N, because $U'>0$ and $M''<0$. Therefore the amount of transmission capacity provided by risk-averse investors is smaller than the amount provider by risk-neutral investors. If investors are risk averse, then they want to limit this downside risk, and will therefore build less generation capacity than in a risk free world.

Summarizing:

> *Proposition 1: In anticipation of aggregate weather-related uncertainty of demand, risk-averse consumers contract for more generation capacity than risk-neutral agents $K_N<K_R$. Risk-averse investors contract for less generation capacity than risk-neutral agents $K_I<K_N$.*

2.3 Exogenous demand uncertainty

Assume we have a homogeneous group of consumers (risk-neutral or risk-averse) exposed to an exogenous demand shock, for instance due to a change of industrial demand. If the consumers have already contracted their expected energy demand, then they can continue with their previous consumption pattern. However, if they change their consumption pattern, then they will only do so to increase their 'wealth'. Energy scarcity will allow the consumers to avoid some of their consumption and sell the energy at a high price to the market. Energy excess allows consumers to obtain additional energy below long run marginal costs on the market. Relative to (2) the 'wealth' function of consumers π not only contains the decision variable K for the investment or long-term contracting decision, but also the option to trade in the spot market by choosing their consumption C different from K. The spot market price $P()$ is an function of non-consumer demand D, consumer demand C, demand shock ε and supply K.

$$\pi(K,C)= M(C) -c^*K+(K-C)P(K-D-C-\varepsilon) \qquad (11)$$

We calculate the first order condition with respect to C to determine the optimal consumption. Note, that individual consumers are not assumed to influence the market price, $P'=0$, which is the typical assumption of competitive markets. The marginal monetary value of consumption equals the spot market price of electricity: $M'(C)= P(K-D-C-\varepsilon)$.

To determine the optimal investment quantity K of risk-neutral consumers under uncertainty we form the expectation of (11) over all ε. With C_ε we express that consumption is chosen as function of the realization of ε.

$$\pi(K)=E[M(C_\varepsilon) - C_\varepsilon P(K-D- C_\varepsilon-\varepsilon)] + K\ \ E[P(K-D- C_\varepsilon-\varepsilon)-c] \qquad (12)$$

Differentiating with respect to K and remembering that consumers do not anticipate their impact on market price $P'=0$ we retain again the equilibrium investment quantity K_N of risk-neutral agents:

$$c=E[P(K_N -D- C_\varepsilon-\varepsilon)] \qquad (13)$$

To determine the equilibrium invested quantity with uncertainty we expand the left hand side as a

Taylor series. Because we have normalized the equilibrium demand such that $E(\varepsilon)=0$, the first order term is zero and the dominant term is of the second order:

$$c = P + \frac{\sigma_\varepsilon^{\;2}}{2} P'' \tag{14}$$

Note that P is a function evaluated at $\varepsilon=0$. If demand is linear, $P''=0$, then the invested quantity is independent of the variance of uncertainty $\sigma^2=E(\varepsilon^2)$. If demand is convex, $P''>0$, then with increasing uncertainty investment will increase such that the market clearing price at the equilibrium is below long-run costs. This corresponds to the previous result that $K_W < K_N$.

Now assume consumer are risk averse, represented by a concave utility function $U'>0$, $U''<0$:

$$\mathrm{U}(K)=E[U\{M(C_\varepsilon) - C_\varepsilon P(K\text{-}D\text{- } C_\varepsilon\text{-}\varepsilon)+K(P(K\text{-}D\text{- } C_\varepsilon\text{-}\varepsilon)\text{-}c)\ \}].$$

Differentiating with respect to K gives (P'=0):

$$c = E\left[\frac{U'\{\mathrm{M}(\mathrm{C}_\varepsilon) - \mathrm{C}_\varepsilon P(K - D + \mathrm{C}_\varepsilon - \varepsilon) + K(P(K - D + \mathrm{C}_\varepsilon - \varepsilon) - c)\}}{E[U'\{\mathrm{M}(\mathrm{C}_\varepsilon) - \mathrm{C}_\varepsilon P(K - D + \mathrm{C}_\varepsilon - \varepsilon) + K(P(K - D + \mathrm{C}_\varepsilon - \varepsilon) - c)\}]} P(K - D + \mathrm{C}_\varepsilon - \varepsilon)\right] \tag{15}$$

To determine the impact of the weighting factor which distinguishes (15) from (13) we expand the nominator and denominator of (15) as Taylor Series in ε. To simplify notation we define $\Delta=K\text{-}C_{\varepsilon=0}$ as the consumer contracted capacity above average demand and obtain:

$$c = P + \frac{\sigma_\varepsilon^{\;2}}{2} \frac{P'' + \Delta \frac{U''}{U'} P' P}{1 - \frac{\sigma_\varepsilon^{\;2}}{2}\left(\frac{U''}{U'}\frac{\partial C_\varepsilon}{\partial \varepsilon} P' - \Delta \frac{U''}{U'} P'' - \Delta^2 \frac{U'''}{U'} P'^2\right)} \tag{16}$$

Note that P, V and C are functions of the equilibrium investment quantity K evaluated at $\varepsilon=0$.

We compare (16) with (14) to see whether risk-averse consumers invest in or contract for more generation capacity than risk-neutral consumers in the presence of uncertainty about third party electricity demand. Subtracting the right hand side of (14) from the right hand side of (16) gives:

$$rhs(16) - rhs(14) = \frac{\sigma_\varepsilon^{\ 2}}{2} \frac{\frac{\sigma_\varepsilon^{\ 2}}{2} \frac{U''}{U'} \frac{\partial C_\varepsilon}{\partial \varepsilon} P' - \Delta \frac{\sigma_\varepsilon^{\ 2}}{2} \frac{U''}{U'} P'' - \Delta^2 \frac{\sigma_\varepsilon^{\ 2}}{2} \frac{U'''}{U'} P'^2 \, P'' + \Delta \frac{U''}{U'} P' P}{1 - \frac{\sigma_\varepsilon^{\ 2}}{2} \left(\frac{U''}{U'} \frac{\partial C_\varepsilon}{\partial \varepsilon} P' - \Delta \frac{U''}{U'} P'' - \Delta^2 \frac{U'''}{U'} P'^2 \right)} \tag{17}$$

If demand is linear *(P''=0)* then (14) shows that risk neutral consumers invest to match expected demand. Assume risk averse consumers also contract to match expected demand (*Δ=0)*. Then the nominator in (17) only contains the first, positive, term (*U''<0* and *C'<0*)[3] and the denominator is also positive. This implies that the rhs of (16) is bigger than the rhs of (14) which equals c. The marginal value of investing or contracting for additional generation capacity exceeds the marginal costs for risk averse consumers. Therefore they will in equilibrium build more generation capacity than (*Δ=0)*.

The argument also applies to small levels of concavity of P', because according to (14) *Δ* is continuous in P''. For larger *P''* the expression does not provide a general result because it is not possible to determine the sign of the nominator of (17) without further assumptions on the convexities.

More interesting is the comparison with risk-averse investors. Investors will again face uncertainty about price, identical to (9). Replacing M'(K-ε) by P(K-ε) and applying the Taylor expansion in ε gives:

$$c = P + \frac{\sigma_\varepsilon^{\ 2}}{2} \left(P'' + \frac{U''}{U'} P'^2 \right) \tag{18}$$

Comparing (18) and (14) we observe that risk-averse investors have an additional, negative component on the rhs. (*U''<0*). Therefore they will in equilibrium invest less than risk-neutral consumers.

As shown above investors without long-term contracts will provide for less generation capacity than risk free agents. With long-term contracts the risk for investors is covered and they provide for similar amounts or even more generation capacity than risk free agents.

---

[3] From (11) follows that consumers choose consumption to equate marginal monetary value and price *M'(C*$_\varepsilon$*)= P(K-D-C*$_\varepsilon$*-*ε*)*. Differentiating with respect to ε gives *C'=-1/(1+ M''/P')*.

Proposition 2: In anticipation of an exogenous demand shock (e.g. to an other country or industry), risk-averse consumers contract for similar amounts or more generation capacity than risk-neutral consumers which in turn contract for more generation capacity than risk-averse investors.

2.4 Uncertainty about individual demand

Third, assume consumers anticipate shocks to their individual demand. We can differentiate between (A) a shift of demand and (B) a shift of available contracted capacity. A shift of demand could be due to a new electric appliance which increases overall utility and is represented by a shift of demand curve. A shift of supply could be either caused by a failure of contracted production or because individual electricity demand increases (for instance because more water has to be heated due to illness or malfunction of appliances).

In both cases, the deviation from expected demand can be traded in the energy market at the energy price which is virtually unaffected by the individual change. A risk-averse consumer would want to limit the downside risk, and would therefore reduce investment relative to the risk-neutral agent if a shift of demand is expected. This is the same shift a risk-averse investor facing uncertain demand would require.

3.3.1 Individual demand shift due to negative events

Consumers realize 'wealth' π as function of consumption pattern they chose and the initial capacity investment they realized.

$$\pi(K,C)= M(C_\varepsilon+\varepsilon) -c^*K+(K-C_\varepsilon)P(K-D- C_\varepsilon) \tag{20}$$

They will adjust C_ε to ensure marginal monetary value of electricity consumption equals P, and therefore $C_\varepsilon = C_0-\varepsilon$.
If consumers are risk neutral, then they will maximize

$$\pi(K)= E[M(C_0) -c^*K+(K-C_0+\varepsilon)P(K-D- C_0+\varepsilon)] \tag{21}$$

As consumers are assumed to ignore their impact on market price (P'=0) the impact of

uncertainty is only a shift of the benefit function, while capacity K is chosen to equate marginal costs of expansion c and expected price P.
If consumers are risk averse, then they maximize:

$$U= E[U\{M(C_0) -c^*K+(K-C_0+\varepsilon)P(K-D- C_0+\varepsilon)\}] \tag{22}$$

Developing the Taylor series second order in ε shows that risk reduces utility:

$$U = U_{\varepsilon=0} + \frac{\sigma_\varepsilon^{\;2}}{2} U_{\varepsilon=0}''PP. \tag{23}$$

Differentiating (23) with respect to K to obtain the optimal capacity choice we obtain the equilibrium condition. (Note, that this time we also include the derivative of P with respect to K, as we are assessing the aggregate impact of consumers.)[4]

$$c = P + \left(1 - \frac{\partial C_0}{\partial K}\right)\left[(K - C_0)P' + \frac{\sigma_\varepsilon^{\;2} U''_{\varepsilon=0} PP'}{U'_{\varepsilon=0} + \frac{\sigma_\varepsilon^{\;2}}{2} U'''_{\varepsilon=0}}\right] \tag{24}$$

First note, that c=P if the group of consumers we are analyzing constitutes the entire electricity demand such that $\partial C_0 / \partial K = 1$. Consumers only ask for additional investment in generation which would allow them to better hedge against uncertainty, if some other user groups would in parallel reduce their investment by some, not necessarily by the same, amount.

Assume consumers would contract for their expected demand $K=C_0$. Then the term at the right hand side in the square brackets is positive (U''<0,P'<0). The marginal value of contracting transmission capacity for consumers is above marginal costs, therefore consumers would increase investment relative to the equilibrium quantity. In this case $(K-C_0)P'$ makes a negative contribution to the right hand side until the equilibrium is reached. The equilibrium invested quantity is above the quantity provided by risk neutral consumers. Increasing equilibrium investment reduces the risk consumers face when they are short of electricity, a benefit risk

---

[4] Note that $\frac{\partial P}{\partial K} = \left(1 - \frac{\partial C_0}{\partial K}\right)P'$ and

$$\frac{\partial U_{\varepsilon=0}}{\partial K} = \left(M'\frac{\partial C_0}{\partial K} - c + \left(1 - \frac{\partial C_0}{\partial K}\right)P + (K - C_0)\left(1 - \frac{\partial C_0}{\partial K}\right)P'\right)U' = \left(-c + P + (K - C_0)\left(1 - \frac{\partial C_0}{\partial K}\right)P'\right)U'$$

The second equality follows because in equilibrium marginal utility M' equals price P.

averse consumers are willing to pay for with higher average costs.

Assume risk averse consumers face the risk that capacity they invested in will not be available. Then they maximize the utility function:

$$U= E[U\{M(C) -c^*K+(K-\varepsilon-C)P(K-D- C)\}].$$

which exactly corresponds to (21) so that we obtain the same outcome.

3.3.2 Individual demand shift due to additional utility from electricity

Assume risk averse consumers face uncertainty as to how much they will value electricity. We represent the outcome by a scaling the utility function with $(1+\varepsilon)$.

$$U= E[U\{(1+\varepsilon)M(C) -c^*K+(K-C)P(K-D- C)\}] \tag{25}$$

In the energy market consumers will continue to equate price and marginal monetary value of electricity $P=(1+\varepsilon)M'(C_\varepsilon)$. Using a first order Taylor expansion we write C_ε as linear function of ε $(C_\varepsilon=C_0+\varepsilon\delta)$:

$$P=(1+\varepsilon)M'(C_\varepsilon) =(1+\varepsilon)M'(C_0+\varepsilon\delta)\sim M'+\varepsilon M'(C_0)+M''(C_0)\ \delta\varepsilon,$$

and therefore $C_\varepsilon= C_0-M'/M''\varepsilon$. Substituting C_ε in (25) gives:

$$U= E[U\{((1+\varepsilon)M(C_0-M'/M''\varepsilon) -c^*K+(K - C_0+M'/M''\varepsilon)P(K-D- C_0+M'/M''\varepsilon)\}]. \tag{26}$$

and developing Taylor series second order gives (P'=0 for the individual reaction):[5]

$$U=U_{\varepsilon=0}+\frac{\sigma_\varepsilon^{\ 2}}{2}\left(\left(\frac{M'}{M''}P\right)^2U''+M^2U''+2P\frac{MM'}{M''}U''-U'\frac{M'^2}{M''}\right). \tag{27}$$

Differentiating with respect to K to obtain the optimal capacity choice, gives analogous to (24):

---

[5] First order: $U'(M-\varepsilon M'M'/M''+PM'/M'')$ Second order: $U''(M +PM'/M'')^2+U'(-M'M'/M'')$

$$c = P + \left(1 - \frac{\partial C_0}{\partial K}\right)\left[(K - C_0)P' + \frac{\sigma_\varepsilon^2\left(\left(\frac{M'}{M''}\right)^2 P + \frac{M'}{M''}M\right)P'U''}{U' + \frac{\sigma_\varepsilon^2}{2}\left(\left(\frac{M'}{M''}\right)^2 P^2U''' + M^2U''' + 2PMU'''\frac{M'}{M''} - U''M'\frac{M'}{M''}\right)}\right]. \tag{28}$$

Comparing with (24) we observe the same structure, with the basic implication that consumers are happy to deviate from expected contracting volume, if that allows them to do risk hedging. In first order this is only possible, if other electricity users alter their contracting volume in reaction such that $\partial C_0 / \partial K < 1$. A small difference between (24) and (28) is, that now the impact of ε is weighted by *M'/M''*. This accommodates for the change that ε is not applied to shift demand but to scale utility. The main difference is the new, second, term in the nominator of (28). As *M''<0* the new term pushes towards less investment. This is because consumers' 'wealth' is increased from the additional services provided by electricity when the monetary value of electricity is increased. This more than compensates for the additional expenditure to acquire more electricity and implies that an upward shift of demand is a positive event which receives less weight by the risk averse consumer. In contrast, a reduction of demand because less electricity services are required, reduces 'wealth' more than the corresponding savings on the electricity bill – a negative shock which receives extra weight by risk averse consumers. To minimize the impact of the negative shock consumers reduce the fixed expenditure on electricity caused by investment or long-term contract at the expense of higher peak prices. This second effect dominates and risk averse consumers invest less than risk neutral consumers, if monetary is extremely convex:

$$\frac{-M''}{M'} > \frac{P}{M} \sim \frac{M'}{M}. \tag{29}$$

3.3.3 Investors' reaction to risk of failing generation plants

Risk averse investors invest less than risk neutral investors if they anticipate uncertainty about the availability of their generation assets. They will enter the market until expected profits are zero:

$$E[U\{P(K_l)(K+\varepsilon)-cK\}]=0 \tag{30}$$

Using, as previously, the Taylor expansion up to the second order and then using a linear approximation of U around 0 gives the equilibrium condition:

$$c = P + \frac{\sigma_\varepsilon^{\ 2}}{2K}\frac{U''}{U'}P^2 . \tag{31}$$

The second term is negative (U’’<0) and therefore in equilibrium the expected market price P exceeds marginal cost of expansion. Investors require a risk premium to cover uncertainty of their production and therefore invest in less generation capacity than risk neutral investors would do.

> *Proposition 3: If risk-averse consumers face individual uncertainty about their effectively available generation capacity, then they will contract for weakly more generation capacity than risk-avers investors.*

All three types of uncertainty lead to insufficient investment in generation capacity in competitive electricity markets, in the absence of a sufficient volume of long-term contracts. So why do we anticipate that consumers will not sign a sufficient volume of long-term contracts?

3 Why consumers will not sign a sufficient volume of long-term contracts

Long-term contracts would cause consumers to reveal their expected future demand for generating capacity to generating companies, as the retail companies (who purchase power on behalf of their customers) would reveal their expected peak demand when negotiating the contracts. This would improve the availability of information to generating companies and therefore reduce their investment risk. Long-term contracts would remove much of the price volatility, which is a risk for generators and consumers alike. An important additional benefit of long-term contracts would be that they reduce the incentive for generators to withhold capacity during periods of scarcity. If generators sell a smaller fraction of their output in the spot market, they benefit less from higher spot prices. So if there are so many benefits to long-term contracts to all involved parties, why are peaking units not covered by long-term contracts in practice? The answer is two-fold:

- With competition among retail companies, most liberalized markets lack a counterpart to sign long-term contracts for final consumers.
- The forward prices for electricity currently only reach up to two years, so it is difficult to anticipate future market prices.
- Some of the benefits of long-term contracts, like reduced exercise of market power, can also be shared by consumers that do not sign long-term contracts.

3.1 Lack of counter-party

Generators do not sell their electricity directly to the final consumers but to retail companies which act as intermediaries. Retail companies, who are the natural counterpart for long-term contracts, possess insufficient credibility as counterparts to the generation companies for long-term contracts. If the average spot price in a given period is below the long-term contracted price, new supply companies will enter the market and sell at the new spot price. Supply companies with long-term contracts will be forced to sell below their costs and make a loss. (This situation will only arise with the total opening of the market when every consumer can freely select supply companies, as implied by the 100% market opening in the proposed EU directive.)

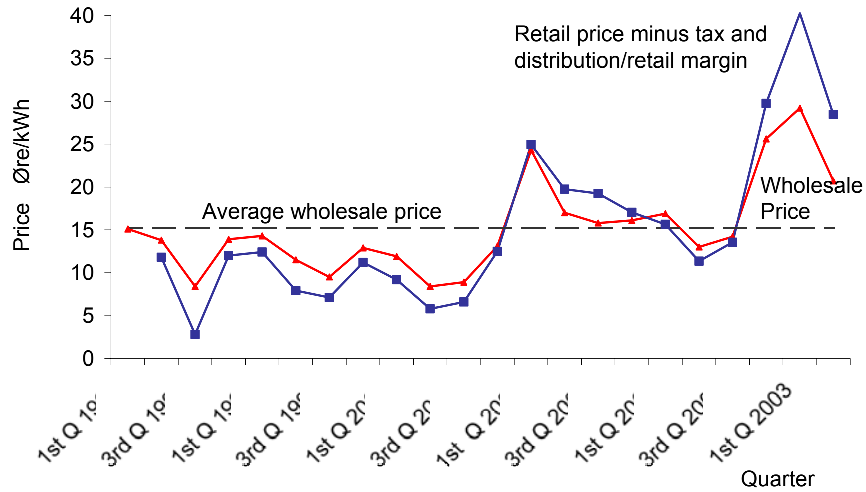

Figure 4: Norwegian retail prices linked to wholesale price.[6]

Figure 4 explains why generators will at the most sign a limited amount of long-term contracts with supply companies. Assume that the average wholesale price during the observation period corresponds to the price of a long-term contract. In periods with average wholesale prices and retail prices above long-term contract prices, like in 2003, supply companies profit and generators lose on their long-term contracts.

In exchange, generators would expect to win from long-term contracts in periods with low wholesale prices, like during 1999-2000. But in such periods new supply companies will offer cheap retail electricity. To retain their customers, all supply-companies would need to follow.

[6] Source: Statistics Norway, http://www.ssb.no, The distribution/retail margin was assumed to be the average difference between retail price excluding tax and wholesale price during the observation period: 7.15Øre/kWh.

Supply companies with existing long-term contracts would incur losses. Some eventually would go bankrupt and would not honor their contracts. Generators would anticipate the resulting decrease in profits from long-term contracts and therefore be reluctant to sign significant volumes of long-term contracts with supply-companies other than for base load, for which the prices are stable.

The risk to generators stems from the fact that supply companies may lose their customers to new supply companies in times when their long-term contracts exceed the short-term price. An institutional change, which would create a credible counterpart for generators to sign long-term contracts, could solve the problem. If, for example, retail companies held regional monopolies, if there were a single buyer or if all consumers purchased their electricity directly from the generators (such as in the system of capacity subscriptions, proposed by Doorman (2000)), consumers would not have the option to switch and therefore the long-term contracts would be credible to the generators.

An exception is when the consumer franchise is retained (Newbery, 2002). With captive consumers, the free-rider problem is solved and generating companies have a credible counter party in the retail companies. This solution would imply that the idea of competition among retail companies is abandoned. Given the small share of costs and of services provided by the retail companies the efficiency losses in a regulated environment should not be too large, while eliminating the option for final customers to switch suppliers would reduce the transaction costs of the required metering and administrative systems.

As an alternative to long-term contracts signed by the local retail company, one could envisage that consumers sign financial contracts for difference with generation companies for their expected electricity demand. Such contracts would not need to be linked to electricity demand or supply but could be a pure risk hedging instrument distributed by banks or any other type of institution. If wholesale prices exceed the long-term average price, then consumers are reimbursed by the generation companies for the difference; if they fall below the long-term average price then consumers pay the difference to the generation company. In a simplified, theoretical perspective this approach would provide the same degree of risk hedging as long-term contracts signed by the retail company. However, the practical implementation might suffer from several factors. If electricity prices are low and generation companies expect money from customers they may face the typical difficulty of a creditor – how to collect the money from individual small customers. Furthermore, in a society that is becoming increasingly mobile, it seems difficult to receive money from customers if they move away, while the same customers might be substantially more willing to stay in contact if they can receive money at times of high wholesale prices. While it seems difficult to implement financial contracts to allow for risk hedging, it seems impossible to achieve a close to complete coverage with these financial contracts. Therefore they can not provide the information about mismatch between future demand and supply which one would hope to obtain if retail companies were to sign long-term contracts on behalf of all their customers.

3.2 Insufficient Incentives for consumers

The second obstacle is that long-term contracts generally are not long enough (Ford, 1999). Long-term contracts would need to extend beyond the current phase of the business cycle to cover at least the next phase in order to dampen the business cycle. The physical inertia of the electricity sector and the close relationship between demand growth and the general economy cause the business cycle of the electricity sector to be long, probably more than five years. Because consumers are not used to this long a time horizon in their energy contracting, the observed contract length is significantly shorter. As a consequence, these contracts do not represent the long-run price of electricity, but reflect the current phase of the electricity demand situation. It seems that electricity wholesale prices are quite high during short periods of time, followed by longer periods with low prices. During these low-price periods, electricity contracts which fix prices for one or two years reflect the low prices of the wholesale market and seem more attractive to consumers than longer contracts, e.g. for 7 years, that would be based on the higher, long run average costs. Therefore private consumers are only likely to sign long term contracts during periods with high wholesale prices, when the 7 year contracts are cheaper than the 1-2 year contracts. But if investment is to be secured with the long-term contracts, then it will only come online with large delay. Moreover, if wholesale prices are high, it is the question whether consumers will find anyone willing to offer long-term contracts based upon the lower long-run average.

A final problem with long-term contracts is the slow learning curve of consumers. The long time it takes to develop new capacity and the long life cycle of generating plant provides a serious obstacle to reaching an efficient equilibrium (Vázquez et al., 2000). If investment signals depend upon consumers entering into long-term contracts, consumers need to learn how to value such contracts. As they would mainly learn through trial and error, this would require repeated periods of shortage and high prices. Due to the length of the electricity ‘business cycle’, consumers have few opportunities to learn how to find attractive contracts. One could argue, that consumers typically will not have much experience with other insurance contracts but will still sign such contracts. However, they cannot learn from observing others, as the electricity price of all consumers is correlated. Secondly, it is difficult to see why consumers should be considered as more 'rational' decision makers when contracting for their energy demand than in the case of retirement schemes, where the state usually provides additional incentives or laws to ensure that consumers participate. Moreover, it is likely that each period of shortages will result in changes of the market rules by the regulator, so that the learning curve would need to be started over.

The result is that consumers will probably never learn to cover all of their future demand with long-term contracts, so that electricity shortages will reoccur time and again and the market will never reach an equilibrium. Even if end consumers or their suppliers have a proper incentive to enter into long-term contracts for peaking capacity, it will take unacceptably long before they would know what their actual (long-term) needs are and how to negotiate these contracts.

4 Additional sources of uncertainty

The previous sections argued that a closed electricity system with perfectly competitive market does not provide investors with an incentive to invest in a socially optimal volume of generating capacity. Now we will show that there are a number of exogenous sources of risk, which further discourage investment. As a result, the real investment equilibrium may be even further removed from the optimum, from the perspective of electricity consumers. Moreover, it may lead to investment cycles, as will be discussed in the next section.

4.1 Imperfect information

Producers lack the information needed for socially optimal investment decisions (Hobbs et al., 2001c; Stoft, 2000). 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.

4.2 Regulatory uncertainty

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:

- Recently, a new Electricity Directive was adopted by the EU (Directive 2003/54/EC), the long-term effects of which will take some time to reveal themselves.
- The European gas market also 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 uncertain. This has a considerable impact on a business plan involving today's state-of-the-art gas-fueled generators.
- In addition, there is uncertainty about future European environmental rules, such as cooling water regulations or the effects of the recently adopted CO_2 emissions trading scheme (EC, 2002).

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 exercise of market power at times close to full capacity utilization complicates the assessment whether high prices are due to market power or due to scarcity. (See also Section 6.) Therefore exercise of market power can induce policy makers and regulators to implement price caps below the value of lost load. Such price caps reduce the expected return of electricity investment and thereby reduce the equilibrium investment volume. 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).

4.3 Regulatory restrictions to investment

Obstacles to obtaining the necessary permits may form another cause of underinvestment. While the social benefits of a proper licensing process are not disputed here, it should be taken into account that it may create negative side-effects. Permits 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. Secondly, permits may be easier to obtain for incumbents, as they already have sites that historically have been used for power generation. This may raise the barrier for new market entrants, which could reduce the incentive for incumbents to expand their volume of generating capacity.

The permitting process can be lengthy, thereby increasing the response time of generation investment to an increase in demand, which may contribute to the development of investment cycles. Especially in a situation of incomplete information about the future development of supply and demand, longer lead times may contribute to investment risk and therefore reduce investors' willingness to invest in anticipation of expected future demand. On the other hand, generating companies may obtain permits well before they actually expect to need them. The (relatively small) cost of the permit process may be considered as the option price for being able to commence construction at any time.

5 Investment cycles

Risk-averse behavior by generating companies, combined with any of the other potential causes of market failure which were described in Section 4, may lead generating companies to delay investment until the demand for new capacity is certain. Due to the long lead time for new generation facilities, however, the new generating capacity might not become available until after a prolonged period of scarce supplies and price spikes. These high prices could prompt an overreaction from investors, leading to an investment cycle.

A year before the California crisis started, Ford (1999) published a paper in which he used a system dynamics model to show that investment in electricity generation facilities is inherently unstable in a system with rules such as then were in force in California. His explanation is that investment is not aimed at dampening the 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 Ford assumes that 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.

Ford's argument is essentially that 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.

Visudhiphan et al. (2001) contend that investment cycles are not inevitable, as long as investors are able to anticipate market developments. However, as we have seen 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. Due to the many uncertainties, the lack of information and the reasons to behave in a risk-averse manner, it therefore is likely that generating companies indeed will have a short time horizon and invest too late.

6 Market power

A significant vulnerability of electricity markets is that generating companies have both ample opportunity and strong incentives to manipulate price spikes, as was demonstrated during the electricity crisis in California (Joskow and Kahn, 2002). When the capacity margin is slim, or when acute shortages already exist, the low price-elasticity of demand means that a small reduction in the supply of electricity leads to steep price increases. Even for companies with a relatively small market share the temptation will be large to withhold generating capacity, for instance by listing generating units as requiring unscheduled maintenance, because it may lead to significant income transfers from consumers to generating companies (CPUC, 2002).

In a world with complete information, the generation company would withhold just so much capacity that the electricity price would reach the value of lost load (which equals consumers' average willingness to pay) or the price cap, while avoiding the need for service interruptions. However, given uncertainty and intertemporal constraints on changing generation output, a generating company cannot always estimate the precise the amount required to bring price to the price cap. Withholding of generation capacity can therefore result in service interruptions.

Stoft (2002) points out that if the price cap is absent or very high, for instance equal to the value of lost load, the increase in profits from withholding can be so high that it becomes attractive even for small generators to withhold generation capacity. The increases in profit which result from withholding are large, while it is difficult to take judicial steps against this behavior (because one would have to prove for each hour during which it occurs which generators were withholding illegally, and not legitimately out of service). The strong incentives to withhold capacity when it is needed most is a fundamental weakness of electricity markets which rely on price spikes to signal the need for investment, even in the absence of other forms of market failure.

While the higher prices should attract more investment, as they represent an opportunity to make more profit, these distorted prices may induce investors and regulators to act in such a way that the effect is reversed. As was already mentioned in Section 4.2, the suspicion that high prices are the result of the abuse of market power, and not caused by a real scarcity power of generating capacity, may be reason not to invest. Moreover, the suspicion of price manipulation may lead to the implementation of a maximum price (or its lowering, if it already exists), the threat of which discourages investment. Therefore, in addition to its detrimental short-term effects, exercise of market power might also reduce investment.

An established oligopoly of large generators may choose a more stable, long term strategy. If generation companies are able to keep prices above marginal costs during normal market conditions, they may opt to overinvest in order to discourage new entry (Newbery, 2002). It is ironic that an unregulated electricity market can only be expected to provide a sufficient level of reliability if it is not competitive. Relying upon an oligopoly to maintain excess generating capacity therefore hardly is an attractive option, because it would undo many of the gains from liberalization. Moreover, as the continued existence of the oligopoly would be uncertain, this is not a reliable option.

Incumbents may also choose a different strategy, namely to use barriers to entry to reduce supply in order to raise average prices. Barriers to entry are common: permitting is likely to be easier at existing locations, where there often is space for an additional unit (e.g. in the place of a dismantled old unit), and because at these sites the cost of a new unit is lower, if the fuel, electricity and cooling infrastructures already exist. In addition, large incumbent firms may obtain the necessary capital more easily, and the risk of imbalances due to generator outages is smaller for generating companies with a large generation portfolio. The latter is not true in a perfectly competitive market, but observation of existing markets teaches that imbalance charges are much higher than competitive levels. Therefore the presence of a stable oligopoly provides no guarantee that there will be sufficient generating capacity. In the UK entry was possible, because low gas prices allowed entry with relatively small Combined Cycle Generators while existing franchises signed long-term contracts for the electric output thereby securing revenue streams to facilitate financing. Currently gas prices are high and competition in supply prevents long-term contracts.

7 Capacity mechanisms

A number of adjustments to the market structure, which we will call capacity mechanisms have been tried or proposed, for the purpose of securing the adequacy of generation resources. A brief overview of the most important ones follows.

7.1 Capacity payments

Payments for installed or available capacity attempt to convert the irregular revenues from price spikes to a more constant revenue stream for generation companies. They have been tried in Spain and several South-American countries and, in a different form, in the former England and Wales Pool. These payments have as a disadvantage that their effect is uncertain: the payments do not necessarily lead to more investment. Instead of fixing the payment level and leaving the investment level to be decided by the generation market, it is more effective to do the reverse. The most promising capacity mechanisms provide a clear signal to the generation market regarding the demand for capacity, but leave it to the market to finance it.

7.2 *Strategic reserve*

Despite the lack of consensus whether there is a problem, several solutions have been developed, some of which have been implemented. One option which often is proposed is a so-called 'mothball reserve', a collection of mothballed old plants which can be returned to service if necessary. A variation is the tendering procedure, which is proposed in the new directive of the EU (Directive 2003/54/EC). The issue is under which conditions this reserve is deployed. If the market is to perform its regular task and invest in generation capacity, it should be able to rely upon periodical price spikes to finance its investment in peaking units. This means that the reserve should only be deployed at a high price, namely a price equal to the value of lost load. This raises two issues. First, it may be politically unsustainable to allow prices to rise this high for any length of time if they can be lowered by deploying the mothball reserve. After all, the reserve will be something of a public facility. The second issue is that the incentive to withhold capacity by market parties will not be eliminated until the deployment price of the mothball reserve is reached. However, if the reserve is to be deployed at any lower price, it will reduce the incentive to invest. This would, in turn, create a need for a larger reserve.

7.3 *Operating reserves pricing*

Another option is for the system operator to provide incentives by contracting operating reserves, something every system operator already needs to do to maintain short-term stability. The system operator may choose to operate this reserve capacity himself, in which case the same problems develop as with a strategic reserve. He may also choose not to dispatch them, but to allow them to return to the market when prices rise high enough. The system operator may do this by limiting his willingness to pay for the reserves (Stoft, 2002). When prices rise during a period of scarcity, the value of selling electricity in the market at one point will exceed the value of selling capacity to the system operator. The system operator's maximum price functions as a de facto price cap for the market, as the owners of the reserve capacity are willing to offer it to the market at this price. This limits the height of price spikes and reduces the incentive to withhold generating capacity. The system operator's purchases of reserve capacity provide the generating companies with a better sense of total demand and provide them with higher off-peak income, which compensates for the reduction of the height of the price spikes. The difficulty with this system is for the system operator to determine the correct maximum price he is willing to pay in relationship to the volume of the reserve. To determine this price, he needs to know the full price-distribution function.

7.4 Capacity requirements

A system of capacity requirements, such as the ICAP system, is used by PJM on the East Coast of the USA. (For an introduction, see for instance Besser et al., 2002.) In this system, large customers or supply companies representing small customers have to by firm capacity to cover their expected peak demand. Capacity can either be provided by generators that are operable within the control area or by out of area producers if corresponding transmission capacity can be secured or by interruptible load. The system ensures that at times of scarce generation capacity generators receive revenue streams in addition to the energy market. The ICAP system so far does not have a sufficiently long time horizon to facilitate investment, but this expansion is considered as part of the standard market design.

7.5 Reliability contracts

A disadvantage of capacity requirements is that they do not provide an incentive to maximize the availability of reserve capacity. An improvement in this respect is provided by reliability contracts, a system of call options which the system operator purchases from the generation companies (Vázquez et al., 2002). When the spot price exceeds the strike price and the options are called, the producers are required to pay the system operator the difference between the spot price and the strike price. Operating power plants are a perfect hedge for the generators: their net income is equal to the strike price. Generation companies who have sold options which are not covered by available generation capacity when the options are called, lose on those options. This provides an incentive to generation companies to sell an option volume which is equal to the available volume of generation capacity which they control. A second advantage is that the generation companies receive an incentive to maximize the availability of their generation units during periods of scarcity. The system operator determines the level of overall generation adequacy by the volume of options which he purchases.

7.6 Capacity subscriptions

Where the previous two systems still contain an element of central coordination, a system of capacity subscriptions leaves all variables to the market (Doorman, 2000). In fact, this system may be considered more market-oriented than a traditional, unregulated electricity market, because it allows consumers to choose their level of generation adequacy. In this system, each customer needs to purchase an electronic fuse, which can limit his electricity use. The fuses are activated by the system operator during periods of scarcity. Customers can choose the size of their fuses. The fuses are sold by generation companies and need to be covered by available generation capacity. Thus the market for fuses indicates the total demand for generation capacity and provides generation companies with fixed revenues to cover their investments. This system turns security of supply into a private good: consumers can choose their own level of generation capacity that they want to have reliably available. The drawback of this system, compared to the previous options, is that it is more elaborate, as it requires the installation of an electronic fuse at each customer.

The last three options promise to be the most effective, as they provide a means for directly influencing the volume of generating capacity. All three provide a mechanism that allows long-term contracts for generating capacity to be signed, also for peaking capacity. Thus they make the demand for generating capacity explicit and reduce the investment risk. Thus they achieve the same goals as a single buyer model would obtain, in this respect.

8 Trade between electricity systems

A different, more practical aspect, of the issue is how to ensure generation adequacy in the presence of significant volumes of trade between systems. In theory, trade between liberalized electricity systems should not change the basic market dynamics. If the involved systems are liberalized in similar ways, trade between them only represents a scale increase. The scale of the system does not impact the question of generation adequacy, as it is addressed in this paper. A benefit of a larger interconnected system is, however, more stability, as the relative impact of individual generators and capacity additions becomes smaller.

In practice, interconnected electricity systems often have quite different market rules, and the rules for using interconnectors are different from the regular transmission access rules within the systems. Therefore the markets which function within the interconnected systems are not fully integrated, but incompletely linked. This has repercussions upon the generation adequacy in the different markets.

In the case of California, for instance, part of the problem was that investment in generation was not only lagging in California itself, but also in neighboring states. There, however, it did not lead to a shortage, but to a reduction of capacity reserves. When the weather suddenly caused a shortage, these states used their own generation resources to meet their own demand first, selling to California only what excess electricity was left. As a result, California, the importing state, bore the full brunt of a crisis, the roots of which were actually spread among a number of states.

In the European Union, a similar scenario is possible. Article 24 of the Directive allows member states 'in the event of a sudden crisis' to take unspecified 'safeguard measures' (Directive 2003/54/EC). This can be interpreted as giving member states the right to curtail exports temporarily in an emergency. While there may be technical reasons for doing so, this means that in the case of a crisis, the European internal market may fall apart into a number of unconnected national markets. With respect to generation adequacy, this is an important issue as each country therefore should have sufficient reserves of its own to guarantee adequate supply under adverse conditions.

Trade between systems with different rules complicates the implementation of a capacity mechanism. During a regional episode of scarcity, systems who have invested in reserve capacity may find that it disappears to neighboring systems with smaller reserve margins. Harmonization of rules clearly is the solution, but may not be feasible in the near term. Countries who wish to implement a capacity mechanism before regional consensus has been reached face the dilemma between waiting for this consensus, which may take too long, or implementing one on their own, which may be more costly and less effective than a regional solution. The discussed capacity mechanisms all appear to have as a limitation that their effectiveness is impaired when they are implemented in the presence of significant import volumes.

9 Conclusions

The theory that there is sufficient incentive for generators in an energy-only market to invest in capacity does not hold in the presence of uncertainty regarding the future availability of generating capacity or fluctuations in demand due to the weather. In the likely case that generating companies and/or consumers are risk averse, the optimal investment level from the perspective of generating companies is below the level consumers would be willing to finance with long-term contracts. The impact are higher electricity prices and possibly more supply interruptions.

The balance between investment and expected returns from price spikes may easily be upset further by a number of factors. Some of these appear inevitable, such as a number of exogenous causes of additional investment risk. Therefore it appears likely that electricity markets that rely on short-term energy revenues will tend to lead to a shortfall of generation capacity over time, possibly resulting in investment cycles.

A second disadvantage of relying upon periodical price spikes to signal the need for investment in generation capacity, is that these price spikes can be manipulated by generation companies.

This dilutes their effectiveness as an investment signal. Moreover, it may result in large transfers of income from consumers to producers and reduces the operational reliability of electricity supply during these price spikes. Several methods have been proposed to stabilize the market and provide better incentives to generation companies and consumers alike. The main effect of these methods is to make the demand for reserve capacity explicit, which reduces the investment risk for generation companies. They reinstate institutions that could sign long-term contracts for energy on behalf of the final consumers. For consumers, the benefits would be increased security of supply and lower price volatility.

References

Ajodhia, V., Hakvoort, R.A. and Van Gemert, M. (2002) “Electricity Outage Cost Valuation: A Survey”, in *Proceedings of CEPSI 2002*, Fukuoka, 2002.

Besser, J.G., Farr, J.G. en Tierney, S.F. (2002) “The Political Economy of Long-Term Generation Adequacy: Why an ICAP Mechanism is Needed as Part of Standard Market Design”, *The Electricity Journal* 15 (7): 53-62.

CPUC (California Public Utilities Commission) (2002) *Report on Wholesale Electric Generation Investigation,* available at www.cpuc.ca.gov/static/announcements/wholesale+generator+report.pdf

Caramanis, M.C. (1982) “Investment decisions and long-term planning under electricity spot pricing”, *IEEE Transactions on Power Apparatus and Systems* 101 (12): 4640-4648.

Caramanis, M.C., Bohn, R.E. & Schweppe, F.C. (1982) “Optimal Spot Pricing: Practice and Theory”, *IEEE Transactions on Power Apparatus and Systems* PAS-101 (9): 3234-3245.

‘Directive 2003/54/EC of the European Parliament and of the Council of 26 June 2003, concerning common rules for the internal market in electricity and repealing Directive 96/92/EC’, *Official Journal of the European Union* 2003, L 176: 37-55.

Doorman, G. (2000) *Peaking Capacity in Restructured Power Systems*, Thesis (Ph.D.), Norwegian University of Science and Technology, Faculty of Electrical Engineering and Telecommunications, Department of Electrical Power Engineering.

EC (Commission of the European Communities) 2002, *Amended proposal for a Directive of the European Parliament and of the Council establishing a scheme for greenhouse gas emission allowance trading COM(2002) 680 final within the Community and amending Council Directive 96/61/EC*, available at http://europa.eu.int/eur-lex/en/com/pdf/2002/com2002_0680en01.pdf.

Ford, A. (1999) “Cycles in competitive electricity markets: a simulation study of the western United States”, *Energy Policy* (27): 637-658.

Hawkins, D. (2001) *The California Report*, Power Point Presentation, California ISO.

Hirst, E. & Hadley, S. (1999) “Generation Adequacy: Who Decides?” *The Electricity Journal* 12 (8): 11-21.

Hobbs, B., Iñón, J. & Kahal, M. 2001a, “Issues concerning ICAP and alternative approaches for power capacity markets”, in *Proceedings of the Market Design 2001 Conference*, Stockholm 7 and 8 June 2001: 7 - 18.

Hobbs, B., Iñón, J. & Stoft, S.E. 2001b, “Installed Capacity Requirements and Price Caps: Oil on the Water, or Fuel on the Fire?” *The Electricity Journal* 14 (6): 23-34.

Hobbs, B.F., Iñón, J. & Kahal, M. 2001c, *A Review of Issues Concerning Electric Power Capacity Markets*, Project report submitted to the Maryland Power Plant Research Program, Maryland Department of Natural Resources.

Jaffe, A.B., & Felder, F.A. 1996, "Should Electricity Markets Have a Capacity Requirement? If So, How Should It Be Priced?" *The Electricity Journal* 9 (10): 52-60.

Joskow, P. & Kahn, E. (2002), "A Quantitative Analysis of Pricing Behavior in California's Wholesale Market During Summer 2000", *The Energy Journal* 23 (4): 1-35, available at econ-www.mit.edu/faculty/pjoskow/.

Liedtke, M. (2000) *State Regulators Cut San Diego Power Rates 3-2 Vote: Utilities Commission Approves Cap On Costs In Effort To Ease Shock Of Deregulation*, Associated Press, Aug. 22.

Newbery, D.M. (2001) *Regulating electricity to ensure efficient competition*, paper presented at the CEPR/ESRC Workshop on The Political Economy of Regulation, London, 1 November.

Newbery, D.M. (2002), *Regulatory Challenges to European Electricity Liberalisation*, Paper presented at the conference Regulatory Reform - Remaining Challenges for Policy Makers, Stockholm, June 10.

Oren, S.S. (2000) "Capacity Payments and Supply Adequacy in Competitive Electricity Markets", in *Proceedings of the VII Symposium of Specialists in Electric Operational and Expansion Planning*, Curitiba (Brasil), May 21 - 26, 2000.

Pålsson, A.-M. (1996) "Does the degree of relative risk aversion vary with household characteristics?", *Journal of Economic Psychology* 17: 771-787.

PJM Interconnection, L.L.C. (2001) *Reliability Agreement Among Load-Serving Entities in the PJM Control Area, First Revised Rate Schedule*, FERC No. 27.

Shuttleworth, G. (1997) "Getting Markets to Clear", *The Electricity Journal* 10 (3): 2.

Stoft, S. (2002) *Power System Economics: Designing Markets for Electricity*, IEEE Press, Piscataway, NJ.

Willis, K.G. & Garrod, G.D. (1997) "Electric Supply Reliability, Estimating the Value of Lost Load", *Energy Policy* 25 (1): 97-103.

Vázquez, C., Rivier, M. and Pérez-Arriaga, I.J. (2000) *On the use of pay-as-bid auctions in California, some criticisms and an alternative proposal*, University Ponteficia Comillas, IIT working paper IIT-00-077A.

Vázquez, C., Rivier, M. and Pérez-Arriaga, I.J. (2002) "A market approach to long-term security of supply", *IEEE Transactions on Power Systems* 17 (2): 349-357.

Visudhiphan, P., Skantze, P. & Ilic, M. (2001) "Dynamic Investment in Electricity Markets and Its Impact on System Reliability", in *Proceedings of the Market Design 2001 Conference*, Stockholm 7and 8 June 2001: 91-110.