Policy uncertainty and supply adequacy in electric power

Brunekreeft, G.; McDaniel, T.

Publication date:
2005

TILBURG UNIVERSITY

TILEC

TILEC Discussion Paper

DP 2005-006

POLICY UNCERTAINTY AND SUPPLY ADEQUACY IN ELECTRIC POWER MARKETS

By

Gert Brunekreeft and Tanga McDaniel

February 2005

ISSN 1572-4042

POLICY UNCERTAINTY AND SUPPLY ADEQUACY IN ELECTRIC POWER MARKETS

Gert Brunekreeft[1]
(TILEC, University of Tilburg, Netherlands)
g.brunekreeft@uvt.nl
&
Tanga McDaniel
(Department of Economics, Appalachian State University, Boone, NC, USA)
mcdanieltm@appstate.edu

05 January 2005
forthcoming in revised and shortened version: *Oxford Review of Economic Policy*

Abstract: This paper identifies key sources of policy uncertainty which can adversely affect investment in the electric power industry, covering both generation and network assets. We focus on asymmetric, systematic, non-diversifiable risk with spin-offs. In generation, a key uncertainty is the inability of governments to refrain from intervention if capacity gets scarce and prices rise. The policy expectations can in fact be self-fulfilling. One issue in network investment is that price-cap regulation is less suited to handling market risk than rate-of-return regulation. Another issue is the apparent inability of regulators to credibly commit to pre-announced policy. With increasing importance of new investment, a case can be made that (used-and-useful) rate-of-return regulation will regain territory in the future. Furthermore, improving checks and balances within the structure of regulation should be a focus of policy and good governance.
JEL classification: L590, L940
Keywords: electricity, generation, regulation, reliability, networks, policy uncertainty

1 Introduction

Supply adequacy in power markets takes on added importance in liberalized markets. Relevant episodes illustrating inadequacies in the electricity supply industry include the California power crises of 2001, which raised questions of whether liberalised power markets set adequate incentives to invest in new assets, and, the recent blackouts in both the US and Europe, which raised concerns about the adequacy of the high voltage transmission networks.

The US-based North American Electric Reliability Council (NERC) defines reliability as: "the degree of performance of the elements of the bulk electric system that results in electricity being delivered to customers within accepted standards and in the amount desired. Reliability may be measured by the frequency, duration, and magnitude of adverse effects on the electric supply (or service to customers)." Further, reliability is divided into two functional

[1] The authors are grateful to Christopher Bliss, Richard Green, Karsten Neuhoff, Hans-Joerg Weiss and an anonymous referee for useful comments and discussion. Gert Brunekreeft gratefully acknowledges research funding from ESSENT nv.; however, the views expressed in this paper are the sole responsibility of the authors.

attributes: adequacy and security. We focus on adequacy which following the UCTE (2003, p. 7): “measures the capability of the power system to supply the load in all the steady states in which the power system may exist *considering standard conditions*.” This reflects the long-term concern whether there is sufficient investment. We use the term supply adequacy to cover both generation and network adequacy.

The discussion on supply adequacy identifies many problem areas. In this paper we focus on policy uncertainty which we apply broadly as the effect on investment incentives as a consequence of uncertain government policy. A narrow definition would be the uncertain effect of economic regulation alone. We attempt to identify key sources of policy uncertainty, to analyse their effects on supply adequacy and to give recommendations for reducing its impact on investment. Concerning new investment in generation assets, the problem seems to be that governments cannot credibly commit to doing nothing if capacity becomes scarce and price soars. Tangential to this issue is the uncertainty created by potential intervention where there is currently no regulation.

As a benchmark we assume that risks which are symmetric, non-systematic, diversifiable and without (future) spin-offs do not affect investment. The benchmark allows us to identify types of uncertainty which do have an effect, namely: 1) asymmetry, affecting expected returns, 2) incompletely-diversifiable risk which, through the CAPM, affects the risk premium as reflected in the cost of capital, and 3) uncertainty creating real option values.

Arguably the most urgent issue at the moment is whether there will be adequate investment in peak generation capacity. Amidst a wide range of arguments, policy uncertainty emerges as a real problem. The argument is essentially that *political reality* dictates that if generation capacity becomes scarce and prices rise political pressure is likely to force intervention by the regulator or responsible government body. The primary effect of this type of uncertainty is asymmetry: while the upside of the price distribution is likely to be capped, the downside is not. Asymmetry has an effect on expected returns (in contrast to the cost of capital) since the distribution of possible prices is truncated. This line of argument leads to the widely discussed ‘hold-up’ problem in risky projects, which also finds application in network regulation, including (new) interconnectors. The typical problem is that a firm invests in sunk assets, after which a regulator, or more generally the state, principally has the possibility to take away (part of) the assets. If the regulator cannot credibly commit to refrain from opportunistically clawing back the sunk assets, firms will anticipate the possibility which will suppress new sunk investment. These effects received strong attention after landmark cases in the US electricity industry, where the pass-through of costs associated with unused nuclear assets was (surprisingly) disallowed (cf. e.g. Joskow, 1989).

In the context of the electric grid, the more narrowly defined regulatory uncertainty is the key problem since market risks (costs and demand) depend on the type of regulation or if there is regulation at all. Here, the overall effect will be on the cost of capital. Wright, et. al. (2003, p. 121) explain carefully that the CAPM approach allows or even makes it necessary to consider only systematic, non-diversifiable risk; i.e. risk that has a non-zero covariance with a market portfolio and which cannot be diversified with other assets. The admittedly extreme approach only considers systematic risk which implies that many types of policy uncertainty simply have negligible effects unless a relation with a market portfolio can be made plausible. The interesting systematic effects include the consequences of market risk on firms' cost of capital under a price-cap, rate-of-return regulation or no regulation.

Lastly, uncertainty around, for example, environmental or nuclear policy creates a real option value if it pays off to wait and see, in which case one would expect delays in investment. The real options model, as developed by Dixit & Pindyck (1994) essentially says that project valuations and thereby the investment decision should include the potential spin-offs on future prospects and/or other projects. Brealey & Myers (2003) distinguish four different real options: the option to expand, the option to wait and learn, the option to shrink or abandon, and the option to vary the mix of production methods. Of these, the second is the most important for our purposes.

The remainder of the paper is organized as follows: sections 2 and 3 examine generation and network adequacy, respectively, and section 4 concludes.

2 Generation adequacy

2.1 Some evidence on generation investment

In this section we will give some empirical indication of the development of generation capacity. The numbers suggest that there may be reason for concern about future adequacy. Moreover there appears to be a remarkable difference between development in the US, where reserve margins have dropped and Europe where reserve margins are stable. In most cases we will adopt the generation reserve margin as an indicator, which is defined as the difference between installed capacity and peak demand as a ratio of installed capacity. It should be noted though that numbers on generation adequacy presented by different institutions may be difficult to compare since indicators like reserve margin, capacity margin and reserve factors are quite often defined in terms of variations. Moreover, installed capacity is not always fully adjusted for capacity that is not (or only partly) available, reserves and planned outages, mothballed and import capacity.

Figure 1: USA reserve margin 1991 - 2002
Source: EIA Electricity Power Annual, 2002 Table 3.2

The OECD (2002, pp. 22/23) indicate past reserve margins for OECD countries between 1985 to 1999. The European reserve margin has been stable at approximately 33%. The Asia Pacific region started with a 35% margin, dropped to 25% but rose again to 35% since mid-1990. Notably, the US reserve margin dropped steadily from 30% to 16% in 1999, which is low. Figure 1 shows the reserve margin across the contiguous US between 1991 and 2002.[2] Reserve margins are at their lowest during the period after plans for deregulation were underway, but before definitive commitments were made (i.e, 1997-1999 or the time between FERC orders 888 and 2000). Some rebounding occurred around the time of FERC's Order 2000. The picture which emerges here is consistent with a wait and see attitude of how the deregulation process would unfold. FERC orders 888 and 2000 deal primarily with open and transparent transmission access and rules for creating RTOs, but within the 1996-2000 period policies on competition at the wholesale and retail levels also began to take shape. To some extent the later increase in margins were due to merchant investments that were later untenable. Joskow (2003) describes this period as a boom-bust cycle in electricity that was not unlike the high tech bubble of the same time. As the bubble burst many proposed projects were abandoned.

[2] We label the graph with "capacity margin" as in the source although the figure depicts "reserve margins" according to our definition. The numbers are calculated as: (capacity resources – net internal demand)/capacity resources.

The end of the stock market bubble, a better understanding by investors of the real economics and market risks associated with building merchant generating plants and trading commodity electricity, trading and accounting improprieties, credit downgrades, refinancing problems, and uncertainties about the future direction of industry restructuring, wholesale market rules, and retail competition, have decimated the merchant generating and trading sector slashing investment and trading activity and dramatically increasing the cost of capacity for new generating plants and very significantly reducing liquidity in forward electricity markets. (Joskow, 2003, p. 22/3).

Forecasts for 2007 and beyond are typically pessimistic but this is at least partly due to a lack of information on new, unknown or un-declared additions. The phenomenon is illustrated by the apparent slowdown of capacity additions further in time in table 1, which gives planned additions in the USA up to 2007.

	Total (MW)	Natural gas additions	% natural gas
2003	71,392	70,063	98.1
2004	40,531	39,445	97.3
2005	37,116	35,077	94.5
2006	37,429	32,150	85.9
2007	13,166	9,346	71.0

Table 1: USA Planned Nameplate Capacity Additions, 2003-2007
Source: EIA, Electric Power Annual, 2002 Table 2.4

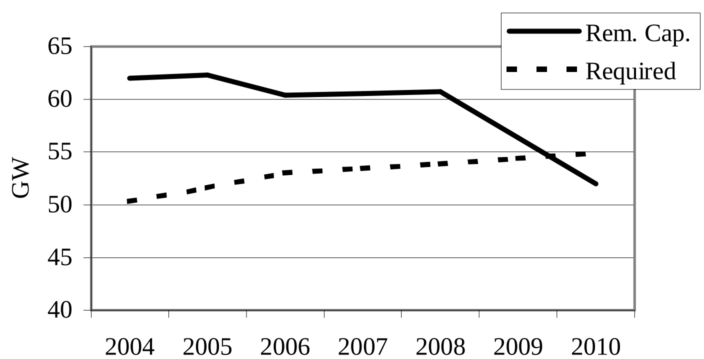

Figure 2: UCTE forecast of remaining capacity.
Source: UCTE (2003)

A similar picture can be seen in the UCTE area if expressed in remaining capacity, as in figure 2. The UCTE covers a large part of Europe except the Nordic countries, Great Britain and Ireland. Remaining capacity is defined as guaranteed capacity less peak load. Guaranteed capacity is installed capacity minus non-usable capacity (including wind) and reserves (some 6% of installed capacity). The UCTE compares remaining capacity with a reserve of 5% of installed capacity, which is considered to be reasonable; this is depicted by the dotted line in figure 2. If remaining capacity falls below the 5% reserve margin, scarcity problems may occur. It should be stressed that this 5% is not short run reserves, as these have been accounted for in the remaining capacity. The 5% reserve margin is a security margin for more serious and persistent disturbances and larger than expected demand. The estimates of UCTE indicate that the margins improve slightly in the next two or three years but fall noticeably after 2007. Again, the post-2007 numbers should be taken with care as they are projections without full information on new and planned investments; at the same time, unknown decommissions are not included either. Remarkably, if the same data are used to make the rather rough calculation of the reserve margin for the UCTE region, we find that this falls slightly from 33% now to 31.5% in 2010. Hence, in these terms one would be rather optimistic, both in time and compared to the USA.

Remaining capacity is typically lower in summer than in winter, which is odd as peak loads (or a proxy of this) are typically in winter-time (mainly February). This can be a combination of two factors. First the problems which have occurred in the UCTE area were predominantly in summer, as the high temperature forced power plants to shut down because cooling water was above allowed temperature. Second, maintenance is typically scheduled in summer (implying that guaranteed capacity is typically lower in summer); one would expect to see more maintenance shifted to milder months.

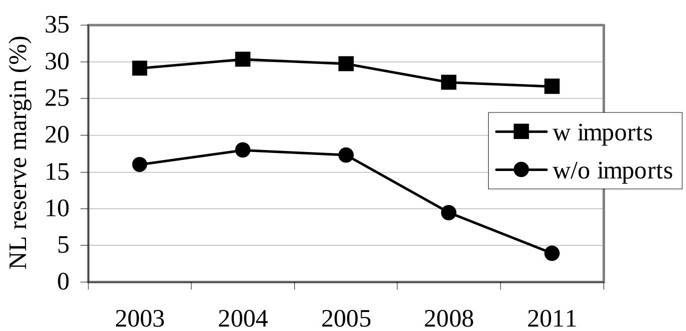

Figure 3: Import dependence in the Netherlands
Source: TenneT (2004)

The Dutch TSO, TenneT, gives a detailed estimate of future capacity margins in the Netherlands (cf. TenneT, June 2004). The resulting reserve margins are given in figure 3. Installed capacity is a conservative estimate as it includes reserves and 20% of non-usable capacity but excludes peaking capacity[3] and capacity for UCTE obligations. Non-usable capacity is low at the moment, but predicted to be 3.3 GW (compared to 19GW peak demand) in 2010. TenneT makes an explicit and important analysis of import capacity, which is reflected in figure 3. Import capacity grows from currently 3.6 GW to a predicted 6.1 GW in 2010 (which is 30% of peak load).[4] This makes all the difference for the reserve margin in the Netherlands. With import capacity, the reserve margin is steady rate at more than 25%. Without import capacity the reserve margin drops to below 4%, which would be too low. Import capacity should be taken into account because it exists. However, there still is a notion that in case of emergencies TSOs may (have to) give priority to their home country and thus governments are getting slightly nervous with import dependency. Furthermore, import capacity is not the same as genuine generation capacity and not all countries can rely on imports at the same time.

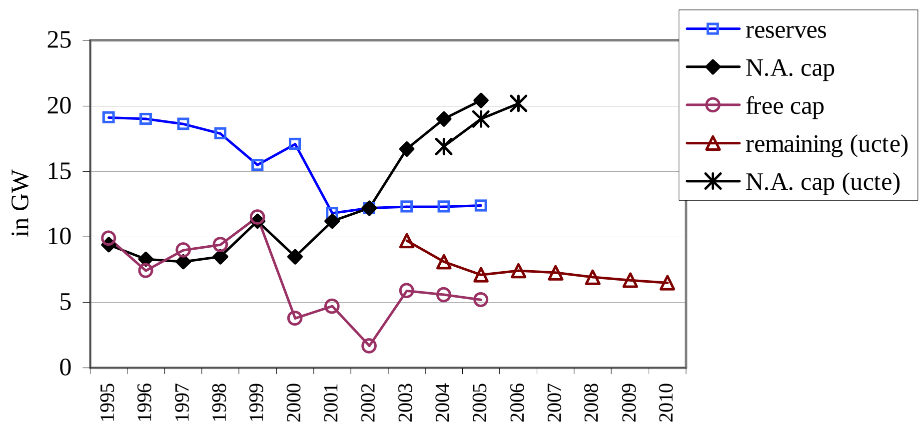

Figure 4: Generation capacity in Germany: the effect of wind
Source: Brunekreeft & Twelemann (2004).

Figure 4 shows capacities in Germany (Brunekreeft & Twelemann, 2004). The calculations combine numbers from different sources which are not perfectly comparable; the main message comes through clearly nonetheless. Following an initial reduction in capacity, installed capacity in Germany grows quite rapidly. It can be argued that the initial reduction of capacity reflects a decrease of excess capacity (e.g., mothballing). The recent increase is mainly wind generation. Wind is expected to grow from about 6 GW in 2000 to over 30 GW

---

[3] Capacity which is expected to run less than 2000 hours.

[4] Apart from steady growth of the interconnectors to Germany and Belgium, this reflects interconnectors to Norway (600MW) and England (1300MW).

in 2008 on a peak load of about 76 GW and installed capacity of about 117 GW. Despite the growth, remaining capacity (as used in UCTE's calculations) falls. The reason is that wind capacity is unreliable as there may be little or too much wind. Load factors are typically somewhat below 20%. This means that most of the capacity growth is non-usable capacity, which enters remaining capacity with a factor 0.2, reflecting the necessity to have back-up conventional assets. Hence, wind turbines may be good for the environment but do little for generation adequacy.

Case studies of the UK and the Nordic countries give an indication of market response to scarcity signals. Generation capacity and subsequent reserve margins in the UK fluctuate quite strongly around its target value of 20%.[5] This ratio is typically net of mothballed capacity. Projections include new planned capacity and decommission as far as they are declared to National Grid Transco (NGT). Taking the broad picture, we find that the reserve margin dropped from over 30% in 1990 to below 20% in 1996, rose steadily to 30% until 2002 and dropped to 20% at the end of 2003. With some reservation, the first decrease may reflect strategic mothballing to increase wholesale prices. As pointed out by Roques et. al. (2004) the first increase was caused by the 'dash for gas'; for a variety of reasons new CCGT entry was profitable and resulted in significant new capacity. It may be recalled that the dash for gas was in fact one of the driving forces behind the new electricity trading arrangements (NETA) which replaced the electricity pool in early 2001. At the introduction of NETA capacity was excessive and post-NETA wholesale prices have been (very) low, resulting in further mothballing and postponement of new projects. This explains the recent decrease of the capacity margin. Notably though, many of these planned projects still have permission.

Roques et. al. (2004, pp. 12 ff.) analyse what they call the first stress test for the UK system.[6] Starting from the 20% reserve margin mid-2003 and following Powergen's announcement to close some 1 GW capacity and mothball another 2 GW, NGT announced in May 2003 that reserve margin might fall as low as 16.2% in the winter 2003/04. Futures prices reacted swiftly raising baseload prices to £33/MWh and £55/MWh for baseload and peakload capacity, respectively. As a result, more than 1 GW mothballed capacity was brought back to the system, restoring the reserve margin. Actual wholesale prices were

[5] Note that in the UK, reserve margin is typically defined as the difference between installed capacity and peak load as a ratio of peak load, instead of as a ratio of installed capacity as by the OECD (cf. JESS, 2003, p. 29).

[6] One of the changes of NETA was to abolish the system of capacity payments which existed under the pool jointly to energy prices. NETA is an energy-only system.

£28/MWh (base) and £42/MWh (peak) in winter 2003/04.[7] Thus it appears that the market handled the first stress test well. It should be noted though that the response is short term and relies on mothballed capacity. There is no indication in this case about new investment. The analysis does suggest though that 'banking of permissions' (i.e., increasing the life span of planning permits) may be an interesting option to enhance adequacy as it increases the speed of constructing new plant.

Capacity margins in the Nordic countries have grown thin. The capacity margin reached its lowest level in 2001 and, as noted by Von der Fehr et. al. (2004), investment has been very low. The authors point to tough environmental constraints as one explanation for low investment. New hydro in Norway is unlikely to get permission, while nuclear in Sweden and Finland is under pressure. Moreover, restrictions are placed on new plants, which emit greenhouse gases. Most notably though is the fact that the authors call the low investment (Von der Fehr et. al., 2004, p. 12) "one of the most important successes of regulatory reforms in the Nordic countries". They first note that the reduction is simply a reduction of excess capacity, and secondly that the wholesale prices and the rate of return on capital in the ESI had been very low (5.5%). Since 2000, the rate of return has been increasing and has almost caught up with the average value of manufacturing industries. Despite political unease, Finland approved a new 1.6 GW nuclear power plant and over 1 GW of new gas plants are planned in Norway.

2.2 Generation investment and policy uncertainty

We treat two types of policy uncertainty affecting generation investment separately. On the one hand, policy which concern primary fuels, and on the other hand, policy to secure adequate generation capacity. For this paper, the latter is more relevant but at the same time less tangible.

2.2.1 Primary fuels

Increasing concern about gas and oil import dependence affects much of energy policy, at least in the EU, thus affecting policy uncertainty for the ESI. Commissioned by the European Commission, the Clingendael Institute (2004) analyses the geopolitical risk in the supply of

[7] Since the actual prices were lower than futures prices, which had predominantly been determined by mothballed and subsequently re-activated plant, one wonders whether traders have been fooled and whether (and how long) gaming by the producers is possible.

primary fuels. The report notes that gas imports will rise to 26% of total gas consumption in 2030 for North America and 63% for Europe.[8] One reason is the increasing share of gas as a primary fuel for electricity production. The concerns for supply security are that strategic dependence on the exporters (the Middle East, Caspian Sea area and Russia) is disturbing, and supply security is threatened by low (indigenous and foreign) investment in the exporting countries.

Obviously, policy uncertainty is beyond the reach of (most) national governments. Yet, the identified problem and policy will affect lower level policies, including: diversification of primary fuels, which covers both fuel sources and importation from different regions; development of an integrated strategic reserves policy (sufficiently large reserves will flatten out short run interruptions); improving demand side management to reduce energy dependency,[9] and developing indigenous energy sources (in particular, renewables like wind and hydro) to reduce dependencies.

January 1st, 2005 was the start of the first round of the European emission trading system (ETS) of tradable CO_2 emission rights. The first "warm-up" round runs until the end of 2007, after which the ETS starts formally running for 5 years (until 2012), thereby implementing the Kyoto protocol. Electricity producers need to have CO_2 emission rights in accordance with production and associated emissions. Rights can be traded and will thus have a price. The effect is an implicit increase in the variable cost of producing electricity which is different for various fuels and technologies depending on their respective emission rates. Consequently, the price of CO_2 rights determines the short term electricity price and production (the merit order) and. More importantly, both timing and technology choice of new investment is affected.[10] The details for the period which just started have long been uncertain. Russia's inclusion in Kyoto will greatly affect futures prices, and their participation has been very uncertain until recently.

The amount of rights allocated to the ESI and the method of allocating rights to new plant (free or auctioned) is important for prices and investment. Helm et.al. (2003) argue that national governments will have difficulty to credibly commit to long-term targets as they will be tempted to renege ex-post. They argue for instance that the UK government did explicitly not commit to binding targets for 2020. As a remedy they favour an independent agency with a duty to reasonably fulfil the targets set by the government. An interesting illustration is the

---

[8] For oil, the figures are 50% and 70% respectively.
[9] Most notably, high costs of energy use will have the long term effect of slowing the growth in energy use. Consequently, energy taxes may have the beneficial side effect of increasing supply security.
[10] See Brunekreeft & Twelemann (2004) for calculations and details.

announced promise of the UK government to increase the number of CO_2 rights beyond its national allocation plan in October 2004.[11] The key uncertainty though is around the details of the new period (from 2008) and in fact the post-2012 period for which it is unclear what will happen. Clearly though, an investment in a large scale power plant built now will run after 2012 as well.

Overall, the CO_2 ETS is a huge achievement and good environmental policy, but its uncertainty will affect investment. Applying the arguments of real option theory, one would expect two effects. First, a slow-down of investment to wait and see. Second, as the relative cost of various technologies is uncertain, one would expect a stronger diversification in different technologies, perhaps combined in one plant but most certainly within firms.

Renewables policy is an additional uncertainty. Renewables are part of environment policy and are affected directly by the CO_2 ETS, but they can also enhance supply security of primary fuels since wind and hydro are indigenous. However, a variety of arguments indicate the limits of the promotion of renewables which would imply a change of the current favourable policy. As such, it seems reasonable to expect some hesitation to further invest.

Lastly, uncertain policy on renewables and nuclear power increases the tendency to wait and see. Both the share of nuclear and the policy towards a nuclear phase-out differ strongly in various states, reflecting controversies. As mentioned above, Sweden, with a 50% share of nuclear has an uncertain policy on nuclear, while (with political unease) Finland recently approved a new nuclear power plant. In 1998, the German parliament decided to phase out nuclear capacity over the course of the next 20 years. The share of nuclear power in German electricity production is 32%. Industry observers expect that this policy will be reversed if the political opposition wins the general elections which are due to take place in 2006. The immediate capacity effect may be moderate as the phase-out program is long and incorporates capacity which would likely be decommissioned anyhow. New investment plans are strongly affected though and are likely to be delayed.

The overall conclusion concerning primary fuels is that policy uncertainty tends to delay investment as a wait-and-see effect. At the same time, this is a type of policy uncertainty which may be inevitable.

[11] As mentioned in the Financial Times of October 28, 2004.

2.2.2 Adequate generation capacity

Stoft (2002, p. 111) points out two demand-side flaws which, at least for the foreseeable future, limit the market's potential to provide sufficient reliability. First, a lack of real-time metering and billing results in little price responsive behaviour, and thus price as a scarcity signal or a rationing device does not work well. Even if consumers were more price responsive, there would be a free-rider problem, which is captured by the second flaw; the system operator's inability to control the real time flow to specific customers. In other words, selective curtailment is not feasible. These two problems together imply that the market might not clear and, if demand is higher than supply, involuntary rationing will be unavoidable. Neuhoff & De Vries (2004) connect this to retail competition and argue that end-user switching impedes long-term contracting, resulting in inefficiently low capacity. Stoft (p. 147) calls the claim that the market will provide adequate reliability a fallacy, since "the system operator must set the price whenever supply fails to intersect demand." Stoft (2002, p. 113) also explains that ISOs in the USA commit to operating reserves of about 10% of load, and pay 'whatever is necessary', and with prices "exceeding a few thousand dollars per megawatt-hour, system operators understandably begin to have second thoughts." As a consequence the ISOs have requested FERC's approval for price caps. As mentioned above these have been set at $1,000/MWh.

At the heart of the matter on adequacy is the rather controversial debate of whether the energy-only market will set sufficient incentives for investment in generation capacity, or whether policy should augment the investment incentives. The energy-only method is, in theory, a hands-off approach whereby the market players should essentially expect the regulator to 'do nothing'. A prevailing view is that energy prices alone (or better yet, energy prices plus short term reserves) will not encourage investment sufficient to meet current reliability standards. The annualised costs of a peaking plant are in an order of magnitude of €40,000 per MW. At a price of €1,000/MWh, the plant would require 40 running hours (cf. similar but slightly higher values for the US in Stoft, 2002, p. 129). Bijvoet et. al. (2003) estimates the Value of Lost Load (VOLL) in the Netherlands on average at a rather high € 8.6/kWh (and even € 16.4 /kWh for households). If the shut down price is set at VOLL of € 8,000/MWh, the running time would only be 5 hours. These prices are very high. Using real data for New England Joskow (2003) shows that energy-only revenues are highly unlikely to recover costs of peaking plant.[12] The robust analysis is somewhat tedious as New England has a price cap of $1,000/MWh on wholesale prices (installed after May 2000), which is considerably below the € 8,000/MWh mentioned above. Still, Joskow shows that the price cap was binding only 6 hours per year on average. He calculates the scarcity rents earned over this period at $10,000/MW-Yr, far less than the cost of reserve capacity which he estimates at approximately $70,000. Perhaps more importantly, as Joskow & Tirole (2004, proposition 5) argue these peak prices are very sensitive to the TSO's actions and incentives.[13]

The Brattle Group (2004, p. 10) concludes: "… we see unique institutional factors that contribute to a fundamental mistrust of markets. For example, the California power crisis has put pressure on most regulators in the United States to intervene directly with security of supply" and "In contrast, the British energy regulator (Ofgem) believes that markets will provide security of supply naturally" and "We believe that Ofgem has been reasonable, and that the distrust of spot markets in the United States may be exaggerated."

This distrust of spot markets takes account of the inevitable political unease with highly cyclical energy prices and the resulting inability of regulators to not intervene during a time of high prices. As an example, consider the case of San Diego just before the California crisis of 2001. Under AB 1890, which laid the foundation for the reformed California market in 1996, retail rates were frozen for six years in order that load serving entities (LSEs) could recoup stranded costs. Companies recovering costs before the 2002 deadline were free to raise rates. San Diego met this criterion in 1999 and doubled retail rates over the period that wholesale prices soared in late 2000. Bushnell & Mansur (2004) discuss this episode and the politicians response; specifically, the passing of Assembly Bill 265 which "froze rates for small and medium-sized (those under 100 kW) retail customers of SDG&E at 6.5 cents/kWh retroactive to June 1, 2000." Ironically, the initial rate freeze was to satisfy producers since the worry was that retail prices would fall in a more competitive environment leaving them unable to recoup existing investment costs. As the San Diego example illustrates, political reality is that regulatory discretion is a huge uncertainty under this alternative. Part of the problem here is poor incentives for demand side participation which is exacerbated by price caps. Energy-only markets provide more volatile prices than other options discussed below (hence their political unpopularity), but without being exposed to the volatility, demand will continue to face few incentives to cut back at times of scarce supply.

The rate freeze in California following deregulation in 1996 is an unusual example (for the utility industry) of a price ceiling that was essentially designed as a revenue floor to ensure that generators recovered stranded costs incurred before deregulation. Following the

---

[12] With annualised costs of $60,000 - $80,000/MWy.

crisis in 2000 the revenue floor became a binding cap illustrating another aspect of uncertainty, namely, the 'stickiness' of policy. Once enacted there can be a significant lag before regulation is revised.

We suggest that policy expectations can be self-fulfilling. Assume that the political reality dictates that the government interferes at some point, and assume further that government cannot commit to refrain from such interference. A strong argument for these assumptions is provided by the OECD, as noted by Roques et. al. (2004, p. 40). Capping high wholesale prices is likely to occur due to the inability to distinguish between the exercise of market power and real scarcity prices. A rational investor has to anticipate and internalise this possibility. If the anticipated interference manifests as an asymmetrical upward cap, expected returns would decrease *ceteris paribus* as compared to a situation where the government could credibly commit to doing nothing. The result is a decline in new investment which leads to scarcity, price spikes and an increased probability of involuntary rationing. This in turn increases the likelihood of an interference. A rational government also has these beliefs. Knowing that it cannot commit to doing nothing, and anticipating rational behaviour by the investors, the government's anticipation of scarcity and the necessity to do something would be self fulfilling. The reverse argument also seems to hold. If both the government and the investors believe that there will be sufficient capacity, the probability of an intervention will be low (or absent), and investors will be willing to invest (more than they would anticipating an interference), and there will be sufficient capacity. It appears that multiple equilibria can exist: low capacity equilibria with interference and high capacity equilibria without interference.

This type of argument is not unfamiliar in the theoretical literature. For instance Cabral (2000, ch.17) describes the existence of multiple fulfilled-expectations equilibria in the context of network effects. The underlying problem is that an individual's decision to adopt a technology with network externalities depends on the expectation whether others will adopt the technology and so on. Armstrong, Cowan & Vickers (1994, p. 188) point out the regulatory problem of circularity, if a regulator determines the regulatory asset base (RAB) relying on the firm's stock exchange value. If the market expects this to be high, the firm's market value will be high and thus RAB will be high and thus allowed revenues will be high and expectation self-fulfil.

---

[13] The resulting uncertainty strengthens the case for using the hedges of Financial Transmission Rights (cf. Brunekreeft, Neuhoff & Newbery, 2004, p. 8).

It can thus be plausible that the high capacity equilibrium will not be reached. This would imply that if the government cannot credibly commit to doing nothing it is the (second) best option to actually do something. The current situation both in the USA and lately in Europe seems to be just this ‘better safe than sorry’ policy. This in line with Oren’s (2004, p. 24) discussion of rationales for capacity mechanisms: “Legitimate concerns for failure of the energy markets to reflect scarcity rents or failure of the capital market to produce proper levels of investment in response to such rents may justify some intervention. In some cases regulatory intervention in adequacy assurance is needed to compensate for regulatory interference in the energy market.”

2.3 Generation adequacy and the search for good policy

Determining an optimal level of capacity (including capacity reserves) must take account of factors such as outside options (e.g., imports), the cost of peaking capacity, and the value of lost load – which will vary internationally and regionally depending on the social cost of outages. This value might be higher than average in Silicon Valley, Brussels or London, for example. The optimal level of capacity trades off a reduction in lost load (greater capacity) and an increase in the cost of meeting load (smaller capacity) (Stoft, 2002, p.137).

There are two diverse opinions about the market’s potential to set incentives for optimal investment. These centre on whether or not plant that is only run for a few hours each year will be able to recover fixed cost. The energy-only view suggests that, if prices are allowed to rise to their true scarcity levels, then yes, peaking plant will be able to recover capital costs and peaking capacity will therefore be provided by the market. Energy prices alone are sufficient to induce adequate generation investment according to this view. Texas is an energy-only market with bilateral contracting between LSEs and producers. LSEs must also contract for ancillary services. Relative to California, Texas has surplus generating capacity (adding 4.3GW of new capacity in 2003) and has added over 850 miles of new transmission lines since 1999. As discussed earlier, England and Wales is now an energy-only market.

The energy-only view can be expanded to include VOLL pricing since efficient bargaining dictates that trading should only occur when willingness to pay exceeds the cost of production. During periods of scarcity the SO essentially purchases power on behalf of consumers; capping the purchase price at VOLL would ensure that he never paid more than consumers’ maximum reservation value. This is economically attractive and theoretically produces optimal investment, but is a high risk means of securing investment since high but intermittent price spikes will yield erratic profit cycles for producers (Stoft, 2002) and possibly invite regulatory intervention.

One reason for unease with the energy-only model is that generation security is often described as a public good with the conclusion that, as with other positive externalities, security will be under-supplied by the market. Joskow and Tirole (2004) use the case of a system collapse, resulting from the outage of a given generator or transmission line, to illustrate that operating reserves can be a public good. This view is embodied in Stoft's second demand-side flaw discussed earlier. Since the SO cannot enforce bilateral contracts between producers and load that rely on the transmission network, individual customers cannot be shut down (or left on line) when there is load shedding. Thus, the consumer cannot pay for the property right to a given level of reliability.

An alternative to an energy-only market is to institute a capacity market. Capacity markets can use either price or quantity methods to set incentives for investment. Price incentives take the form of capacity payments such as accompanied the former power pool in England and Wales. With capacity payments, generators declared available during a dispatch time period receive a payment (even if they are not dispatched). Argentina, Columbia and Spain use capacity payments. Even though capacity margins remained high throughout the life of the Pool, the capacity payment was widely criticized for being open to manipulation. This method is transparent, however, and uncertainties surrounding market power can be minimized if there are outside options to deal with anti-competitive behaviour. That was the case in England and Wales since the regulator could refer companies to the Monopolies and Merger Commission (now the Competition Commission). In the U.S., market monitoring units accompany newly established RTOs and these units can monitor and investigate potential manipulation of market rules. Serious allegations of abuse can be referred to the Department of Justice. Arguments against capacity payments include Joskow and Tirole's (2004) illustration that capacity payments will not ensure that peaking capacity covers its cost if market power can affect multiple prices (e.g., peak prices and price close to peak), and Stoft's (2002) discussion of risk and market power as two 'side-effects' of reliability policy. Many of the options for securing reliability, particularly those relying on VOLL pricing, have high but uncertain, infrequent price spikes. This is risky for investors, and high price spikes invite market power abuse.

Quantity methods to procure reserve capacity are either centralized or decentralized, the former involving contracts between the system operator (SO) and producers and/or load, and the latter involving contracts between producers and load.[14] Centralized methods require determination of both a level of "reserve" capacity and a price to pay for reserves.[15] Either or both of these components may be decided by the SO who may not have full information. One advantage of quantity alternatives relative to pricing alternatives is that the SO has some control over the magnitude and duration of price spikes. As discussed in Stoft (2002), the social/political acceptability of price spikes depends on their size and duration. A mechanism that produces low spikes more often has three benefits: (i) it is likely to be more acceptable, (ii) it would be less risky for investors, (iii) it would be less susceptible to market power.

Two theoretically similar centralized methods of securing generation are strategic reserve pricing and operating reserve pricing. In each case the SO determines the amount of capacity to be held in reserve leaving the market to provide sufficient capacity for the energy market. A difference occurs if we interpret the market for strategic reserves as an 'emergency' market, possibly supplied using sources owned by the SO.[16] An emergency market operates independently from the energy market, and could be subject to different, higher price caps than the energy market. According to Stoft (2002) the addition of an emergency market with its own price cap can control the exercise of market power by confining price spikes to this rather small portion of the market. For longer term adequacy this mechanism is less attractive than operating reserve pricing – one reason being that extensive use of an emergency market leaves much discretion to the SO in determining when to call on emergency reserves.

Under reserve capacity pricing the SO has two instruments at his disposal: a maximum willingness to pay for reserves and the level of reserve capacity. The price the SO has to pay for reserve capacity is inversely related to the size of the reserve level. This holds for two reasons: (i) the more power contracted as reserve, the less there is available to be bid in the energy market and (ii) the SO can always reduce the reserve requirement (up to available capacity). The opportunity cost to a generator of contracting with the SO is the price he could obtain in the energy market. If the gap between demand and available capacity is large, the opportunity cost is small, so the SO can maintain a large reserve without paying much for it. As the gap between demand and available supply gets smaller, the SO can release reserve capacity into the energy market (i.e., reduce the size of the reserve) whilst maintaining the price he is willing to pay. At some point, however, as the capacity gap closes, the opportunity cost for generators rises above the SO's willingness to pay, and the SO can only attract more

[14] Our explanation of quantity alternatives follows closely to Stoft (2002) and de Vries (2004).
[15] Here we use 'reserve' to refer to any capacity that is taken out of the energy market and set aside for another purpose such as meeting security requirements.
[16] Conditional on these resources only being deployed as a last resort when shortages occur. (de Vries, 2004).

reserves by paying more for them. The price in periods when reserves are dispatched would optimally be set at the cost of the marginal generator who is not included as part of the reserve. The SO can make this price higher (lower) by decreasing (increasing) the reserve level. A lower reserve level increases the duration of price spikes necessary for new investments.

The method of reserve pricing can be augmented in a number of ways. The SO can contract for capacity on a daily or longer term basis. Longer term contracting would strengthen the price signals to investors and lessen arbitrariness and discretion in setting the reserve capacity level. The Dutch Ministry of Economic Affairs has recently recommended that the SO, TenneT, make broader use of (auctioning) long term contracts to secure adequacy. The Brattle Group study of 2004 recommended focusing contracts on peaking plant since a large part of peaking capacity in the Netherlands is likely to retire over the next 5 years; for the longer term reserve capacity, quick response spinning reserves normally provided by base load are not required and slow response capacity provided by peak plant is cheaper. Market friendly extensions of reserve capacity pricing discussed by Vázquez et. al (2002) and Oren (2004) would allow the use of (call) options to procure reserve capacity.

The US capacity markets aimed at securing long-term adequacy are decentralized requiring individual 'load serving entities' (LSEs) to obtain resources to meet their obligations (usually through bilateral contracts with generators).[17] The additional market for their capacity should induce generators to invest in capacity beyond that required to meet peak load, and because contractual obligations can be traded in secondary markets price discovery is improved. ICAP markets are used in the PJM, NY and New England markets. In PJM, for example, capacity obligations are set such that the system as a whole maintains a loss of load expectation of one day in ten years. Under this mechanism generally, LSEs pay a penalty if they have not contacted with generators for the required surplus power or met shortfalls through interruptible contracts with consumers. The penalty becomes the indifference price between contracting for surplus power or being short and so sets the price for capacity contracts when they are scarce (de Vries, 2004). If the length of the contract period is too short generators will renege on their contractual obligations during times when the spot price is above the penalty price. Requiring yearly contracts and annualizing the penalty surpasses this problem, however (Stoft, 2002). Vázquez et. al (2002) criticize capacity obligations because only the price is determined by competition while the quantity that

[17] ICAP requirements in the US are approximately 118% of peak load (Stoft 2002).

producers are allowed to sell is administered – a problem if there is a large amount of non-thermal generation on the system.

In the search for good policy on generation adequacy we maintain that until technologies for demand side participation become cost effective and private property rights for reliability can be credibly honoured, caution over generation adequacy is apposite and use of market friendly capacity mechanisms can reduce investor risk while appeasing politicians by ensuring a consistent revenue stream for producers without extreme price spikes. Mechanisms which rely on quantity targets may be less prone to market power abuse and have the advantage of allowing for tradable contracts. Centralized mechanisms, such as reserve market pricing can achieve the same result as decentralized methods but at less risk to LSEs who are reluctant to sign long term bilateral contracts with producers when retail competition is active.

Relative to energy-only markets capacity mechanisms reduce the level but increase the duration of price spikes giving generators a more consistent revenue stream and lowering investment risk. Moreover, in an energy-only market the price signal for new investment occurs as *demand* approaches *available supply*. With capacity targets this price signal occurs much sooner since the market becomes tight as *demand plus target* approaches *available supply*, so the investment signal arises in advance of capacity need. The size of the target will determine how soon scarcity is reached and thus the timing of the investment signal.

Capping prices at VOLL might theoretically lead to the optimal amount of capacity being supplied to the market, but VOLL pricing can lead to extreme price spikes which invite market power abuses and create a risky investment environment. Taking political reality and potentially risk averse investors into consideration capacity mechanisms appear to be a better alternative.

3 Network adequacy

3.1 Some evidence on network adequacy

Whereas there is justified concern on generation investment and adequacy, both the theory and the evidence on network adequacy are less clear. Figure 5 gives network capacity over peak demand for the USA and UCTE (which here includes the CENTREL countries). These data give an impression of the capacity of the grid, but in many instances networks are ageing and need substantial investment to be modernised.

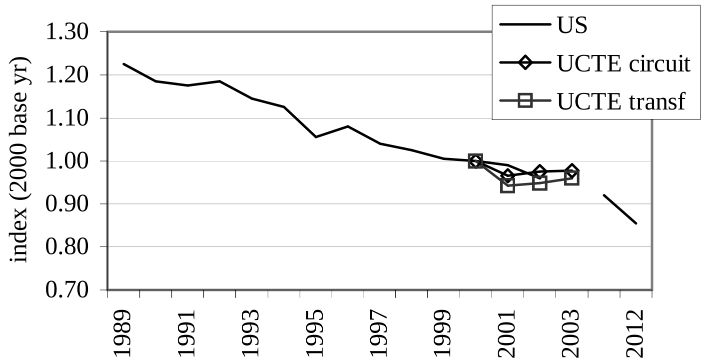

Figure 5: Network capacity in the USA and UCTE
Note: capacity over the system's peak load. Indexed with 2000 as base year.
Sources: Hirst (2004) and UCTE statistics (various years)

The data have been indexed (with 2000 as the base year) to enhance comparability. For the comparable years the developments appear similar, in both cases suggesting a modest decline of capacity. More striking is the big downward trend in the USA. However, it is not a priori clear what this trend tells us. Hirst (2004) notes that this has been a trend since the early 1980's and thus it is unlikely to be the (sole) effect of liberalisation. First, an existing grid can be optimised in use and design, facilitating higher network load. Second, planned grid expansions are often delayed by environmental permission obstacles. Third, since the 1980's power plants have become smaller. CCGT, wind and other renewables are relatively small; distributed generation is small, close to the users, and mostly embedded in the distribution network. The smaller size of power plants means that the grid will take a different shape. One should expect that with decentralisation of power generation, the capacity of the high voltage grid can be reduced *ceteris paribus*. Fourth, generation and transmission capacity can be traded off to some extent. If a line connecting A and B is congested, then the line's capacity might be increased to facilitate the power flow, or new generation capacity can be build on the importing side to reduce the power flow. Fifth, the data given above do not (accurately) reflect the spatial and temporal diversification of the grid and the subsequent load. Finally, grid capacity did grow in the last two decades, but peak load grew faster.

Other arguments for a slow down in grid growth are directly or indirectly related to liberalisation and regulation. There are reasons to believe that for a long time reduced capacity margins reflected a reduction in excess capacity. It is persistently claimed that gold plating may have resulted in excess capacity with inefficiently high reliability. Also, demand growth up to the late 1970s and early 1980s had been large but fell steeply after that. The sudden decrease in demand growth exaggerated excess capacity. Joskow (1989) argues this for generation reserve margins, but one would expect the same to be true for network capacity. Related to this is the argument that regulators disallowed full pass through of capital cost of excess capacity, applying the used-and-useful clause. Lyon & Mayo (2000) suggest that this may have frustrated new investment to some extent.

Furthermore, in both the USA and Europe, trading over networks has increased substantially. UCTE data suggest that cross-border exchanges in the UCTE area[18] increased from about 200TWh in 1999 to 300TWh in 2003. The European Commission (CEC, 2002) outlines the current congestion on European interconnectors. Of 48 two-way cross-border interconnectors, 12 are frequently or always congested and 21 occasionally or seldom. Bialek (2004) argues that increased trade and reliance on the existing networks causes many of the current reliability problems. The interconnectors have simply not been built for extensive trade and time is needed to resolve the scarcity. Also, it is not always cost effective to reduce capacity by expanding interconnector capacity as it might be better to build new power plants.

There is also concern that the price-cap regulation impedes new investment. The arguments are threefold: the first argument is more of an accounting argument than a genuine problem. If depreciation is such that the regulated asset base (RAB) plus depreciation decreases in time (eg. linearly), a significant new investment will require higher charges and hence a revision of the price cap. This reflects a common problem under price capping which at least delays significant new investment. The second concern is that price-cap regulation worsens the ability of regulators to commit to pre-announced policies. This issue will be dealt with in more detail further below. Third, there is justified concern that price-cap regulation impedes quality of supply (QoS). As QoS is a good indicator of network adequacy (and reliability) this is clearly an important point.

Regarding quality, the arguments are relatively straightforward for the short run. Assuming that in the short run, demand is relatively inelastic to quality changes, then decreasing maintenance lowers costs will not (or only hardly) reduce quantity and vice versa. The assessment is unclear for the long run. Demand will adjust to reflect lower willingness to pay due to lower quality. This in turn translates into lower revenues which may offset the cost savings from reduced spending on maintenance. Moreover, in the long run, the price cap will be adjusted to reflect the reduced costs as well (a similarity to rate of return regulation). As long as price is fixed, quality will be lower than optimal, but this does not always hold for flexible prices (Spence, 1975). Because the classical form of rate-of-return regulation induces

[18] Including new members and including exchange with third countries.

excessive quality, one might expect quality to go down in many cases as a result of implementing price-cap regulation.

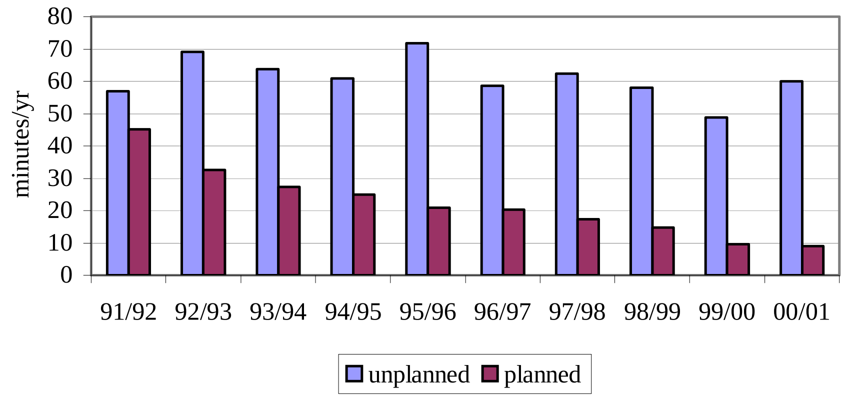

Figure 6: Duration of supply interruptions in the UK
Source: CPB (2004, p. 77).

It is common to distinguish the following quality indicators: frequency of interruptions, duration of interruptions and energy not supplied (ENS). CPB (2004) studied the UK and Norway. Figure 6 shows the development for the price-capped electricity networks in the UK since liberalisation expressed in duration of supply interruptions. There is no clear pattern in unplanned outages but planned outages have gone down. The first is good news and suggests that the quality degradation under price caps may be moderate. The reduction in planned outages may be caused by reduced maintenance, and thus the associated short-run gain in quality may be bought at the expense of future lower quality. This explanation is convincing but one expects to see it manifest in higher unplanned outages at some point. Perhaps it is too early to see this effect, perhaps also the reduction in planned outages is the result of better managed maintenance. The data from Norway add an interesting feature. Under the price-cap regulation ENS improved. This could be because ENS is lost revenues[19] and thus the firms will have an incentive at least not to lose too much power but it might also be the effect of lowering planned outage. CPB (2004, p. 75) suggests that the latter effect is more important.

[19] Also for the network operator to the extent that charges are per kW or kWh.

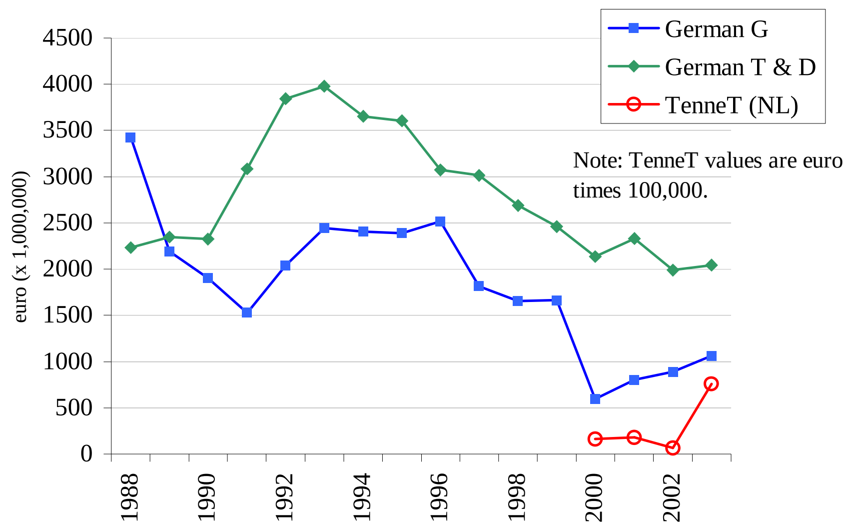

Figure 7: Investment in Germany (aggregate) and Netherlands (TenneT only). Sources: For Germany, Brunekreeft & Twelemann (2004); TenneT's annual reports

A last indication of the network adequacy is the investment activity. Under the EU-TEN programme, 23,000 MW of additional transmission capacity inside or to the EU is planned. Figure 7 gives investment data for the electricity sector in Germany and for the TSO in the Netherlands (TenneT). The trend in Germany is blurred by the re-unification in 1990 which triggered large-scale modernisation of the former eastern ESI. It is re-assuring to see that the 'downward' trend seems to have stabilised. The picture in the UK looks the same (cf. JESS, 2003, p. 37). In the Netherlands, TenneT's investment looks rather optimistic and probably it is. TenneT has announced and started major projects (basically strengthening the high voltage grid in the urban west and upgrading the interconnectors), but at the same time it should be noted that 2003 is the beginning of the new regulatory period and hence the natural period to book significant new investment; the RAB has increased accordingly. Moreover, TenneT is buying up regional high-voltage networks, which increases RAB without actual investment. In the Nordic countries investment is low, but according to Von der Fehr, et. al (2004, p. 17), there is no lack of initiative, but "it would seem that regulatory and political will, rather than commercial will, is going to be decisive."

3.2 Regulation and the cost of capital.

Regulation has an effect on the incidence of market risk on the regulated firm. The reference is the 'buffering hypothesis' put forth by Peltzman (1976, p. 230): "Regulation should reduce conventional measures of owner risk. By buffering the firm against demand and cost changes, the variability of profits (and stock prices) should be lower than otherwise. To the extent that the cost and demand changes are economy-wide, regulation should reduce systematic as well as diversifiable risk."[20] The crucial factor is how much of the shocks can be passed through to customers. Profit-maximising prices of a firm with market power pass through only part of the demand and cost shocks and absorb the other part to restore optimality and thereby profits vary with demand and cost shocks. In contrast, under an admittedly extreme notion of rate-of-return regulation (i.e. with a regulatory gap of zero), the firm in fact passes through all of the shocks in order to stick to the allowed rate of return. Hence, variability in profits is low. The investment in the rate-of-return regulated firm may give low returns but at least they are safe.

The perspective changes for firms regulated with a price cap. Wright et. al. (2003) examine the case of price caps in detail. They conclude that the cost of capital are higher under price-cap regulation than for an unregulated firm for cost uncertainty.[21] This is intuitive because if costs change and prices stay as they are, variability in revenues and profits is strong. For demand uncertainty, cost of capital are lower under price-cap regulation than without regulation. For demand uncertainty, the price cap works as the buffer under rate-of-return regulation. The fact that the firm is not allowed to increase the price (to adjust to demand increase) and, with a binding cap, does not wish to decrease the price (to adjust to a demand decrease) means that the changes in profits are less under the price cap than they would be if the firm would freely adjust to demand changes.

In practice rate-of-return and price-cap regulation are not that different. As has been explained by Joskow (1989), the US procedure in rate hearings is triggered by either the firm or the regulatory commission. This implies that as long as prices fall within reasonable range, nothing will happen and the endogenous regulatory lag can in fact be quite long.[22] Grout & Zalewska (2003) present an interesting study on the effect of different regulatory regimes on risk as measured by CAPM. They study the effect of profit-sharing regimes in the UK during the second half of the 1990s. Profit sharing should here be seen as an explicit modification of price-cap regulation in order to re-allocate high profits made by the firms under the price-cap regulation. Grout & Zalewska (2003) define profit sharing as a weighted average of the outcome under rate-of-return regulation and the outcome under price-cap regulation. The

---

[20] Brennan & Schwartz (1982) confirm the claim with a formal and numerical approach.
[21] Except for the case of complete cost-pass-through.
[22] In fact this may well be asymmetric; nothing changes downward, while upward regulated prices can be adjusted.

weight is the profit-sharing factor. The econometric analysis confirms that profit sharing lowers the risk and thus the cost of capital.

The results suggest that as underinvestment becomes a more serious threat, rate-of-return regulation will become more attractive *ceteris paribus*. In less formal terms, what one would expect to observe is an increasing tendency to shorten and/or endogenize the regulatory lag. Modifying the price-cap rules with an explicit profit-sharing rule may be an interesting hybrid.[23] The task is to design a profit-sharing rule which reduces risk while at the same time retains the virtues of a price-cap rule.[24]

3.3 Asymmetry and dynamic consistency

The problem addressed in this subsection concerns the inability of regulators to credibly commit to an implicitly or explicitly agreed rate of return. The policy uncertainty here is uncertainty around the regulator's behaviour. The effect is asymmetrical capping of revenues and a consequent decrease in expected returns. The cost of capital might be unaffected or at least play only a minor role.

Underlying the discussion is how to achieve consistency of the regulated rate of return. A seminal contribution stems from Myers (1972, p. 79), who studies the 'fair' rate of return in regulation. Myers argues in favour of a competitive benchmark, implying the use of a 'conscious' regulatory lag, thereby effectively arguing for a price-cap regulation *avant la lettre*. Myers' emphasis is on making the concept of a fair rate of return operational and on short-run efficiency effects.[25] Myers also notes the consequence that the risk for the firm might be higher than under the no-lag rate-of-return regulation and the fact that ex-ante price capping might allow (excessively) high profits.

The electric power industry in the USA provides a critical illustration of regulatory uncertainty, which may or may not be opportunism. The power industry in the USA has long been regulated by rate-of-return regulation. However, as explained well by Joskow (1989), rate-of-return regulation in practice meant setting allowed price (derived from allowed profit rate) until a rate hearing would revise these. These rather costly and time-consuming rate hearing could either be called by the regulator or the firm. This principle establishes an

[23] Note however that here the profit-sharing rule would not to share 'excessive' profits, but rather to reduce risk and thereby not impede investment.
[24] An example might be the incentive scheme for the balancing services for the system operator NGT in the UK. The scheme works with sharing factors (up- and downside) and caps and collars around target levels.
[25] Note that the traditional rate of regulation in the USA did have a regulatory lag but that this has not been consciously but rather ad hoc (cf. Joskow, 1974 and 1989).

endogenous regulatory gap and causes prices to be sticky (especially upward). In the 1970s, fuel prices rose and demand growth stagnated, which in turn triggered rate hearings. Joskow (1989) argues that the increasing energy prices put the regulators under pressure to disallow the price rises. In effect, the used-and-useful clause was applied ex post to especially nuclear power plants. The used-and-useful clause implies that costs can only be passed through if the economic value is higher than accounting costs. Lyon & Mayo (2000) estimate that some US$ 19 billion was not allowed. They also note that the bulk of these disallowances was for management imprudence, but major disallowances concerned excess capacity. To our view, the distinction is crucially important. The former is exactly what one would expect a regulator to do and corresponds to regulatory consistency. The latter was opportunistic behaviour and breach of the regulatory compact. Joskow (1989, p. 161) notes critically that: "utilities learned that if they built large new generating plants, they might very well not recover their investment: commissions might resist large rate increases even if the increases were fully justified."

Lyon & Mayo (2000) make an empirical assessment of the disallowances of excess nuclear assets. They find that the propensity to invest for firms being subjected to a disallowance did decline and moderately for nuclear investment by other utilities in the same state. This characterises the non-commitment problem: if the regulator cannot credibly commit to its announced policy, it will frustrate investment.[26] Stated with some reservation, the analysis in Lyon & Mayo (2000) suggests that a used-and-useful rule applied against bad management does not spill-over and is regulatory consistent; at the same time, if it is a breach of regulatory contract it increases uncertainty (asymmetrically) and adversely affects investment. Whether an application of the used-and-useful clause is consistent or opportunistic is ultimately for the investors to decide.

Gilbert & Newbery (1994) offer an interesting perspective to the used-and-useful rule under rate-of-return regulation (UUROR). In essence, the used-and-useful clause may strike a good balance between a pure rate-of-return and a price-cap regulation. Under pure rate-of-return regulation the regulator has been constrained constitutionally to allow a fair rate of return. This may be safe to the firm but it is well known that it induces overinvestment. Price-cap regulation can be seen as a regulatory system in which the regulator is not

[26] They also find that the reputational spill-over effects on non-nuclear investment is small or even reverse, which is counterintuitive. The argument may be that the spill-over effect (which reduces investment) is compensated by better investment opportunities in non-nuclear assets. Lyon & Mayo (2000, p. 8) also conclude that the isolated nature of the reputational spill-over may hold for transmission and distribution facilities. However, as the isolation effect claims that it does not spill-over from one utility to the next, it remains unclear whether the spill-over is isolated form one department to the other in the same utility.

(constitutionally) constrained to allow a fair rate of return. In this case, regulators may be tempted to deviate from previously announced policies and will have difficulty to commit to not doing so. Anticipating this, firms will tend to underinvest. By allowing to shave off extreme and unreasonable outcomes, a UUROR decreases the regulator's incentives to cheat and thereby increases the regulator's credibility. In more formal terms, in a repeated game, the UUROR enhances the set of subgame perfect Nash equilibria along the optimal investment path. At the same time, proper application of the UU clause mitigates the goldplating incentives under the pure ROR regulation.

The asymmetry policy uncertainty, which has been addressed above, played a role in a recent Australian debate (cf. e.g. Gans & King, 2003), in particular, with respect to new risky interconnectors. The discussion appears to have found its way in the EU Regulation on Cross-Border Exchange of July 2004. In Australia this was one of the arguments to allow unregulated merchant investor for interconnectors.[27] A handful of such projects have been completed with mixed success. The European approach is laid down in the EU Regulation on Cross-border Exchanges which entered into force July 1st, 2004.[28] Art. 7 of the regulation allows for *new* interconnectors to be exempted from art. 6(6) of the regulation and arts. 20 and 23 of the EC electricity directive.[29] The former specifies regulation of the revenues of allocation of scarce interconnector capacity, while the latter requires (regulated) third party access to the network. One of the conditions to qualify is that the project is risky (cf. Brunekreeft, 2004). In Europe no Art. 7 projects are known at the moment of writing. In the USA, some merchant projects into the New York area have been proposed, but the investors have difficulty to find funds.

3.4 Good governance

The policy discussion can be captured by the term *good (market) governance*. Hancher, Larouche and Lavrijssen (2003, p. 356) define good governance as: "the search for the best set of all laws, regulations, processes and practices that affect the functioning of a regulatory framework and the market". Amidst a long list of aspects (cf. Hancher, et. al., 2003), we think that two aspects stand out for the regulation of the complex network industries: independence and flexible powers.

---

[27] To be precise, for underinvestment, the expected rate of return should be below required cost of capital. This is the case if the benchmark is regulation with a fair rate of return, but (need) not if the benchmark is no regulation.
[28] Regulation 26 June 2003 (1228/2003) (1/7/04).
[29] EC Electricity Directive 2003/54/EC, 26 June 2003.

Independence has been well formulated by Ocana (2003, p. 17) as meaning "that the regulatory agency is protected from short-term political interference." Indeed, a regulator with the authority to create and divide rents will be subject to capture by politics. The politician is likely to have different aims than the regulator. Politicians, subject to re-elections, will tend to have a higher 'political discount factor', and will therefore have a stronger tendency to 'make hay while the sun shines'.[30] The effect of political interference can go two ways depending on ownership. If the state is owner of the regulated industry it will have an interest in high prices and subsequent profits. On the other hand if the regulated industry is fully private, the state may follow consumer interests and pursue low prices. In that case, the industry runs into the hold-up problem as outlined above. Spiller (1996 quoted in Ocana, 2003, p. 17) points out that statutory independence of a regulator can be seen as a commitment against opportunistic behaviour. Changing legislation is costly and time-consuming and usually requires parliamental approval.

Independence of regulators requires careful restriction of regulators' authority. This in turn raises a problem. Too much detail is likely to become unworkable for complex and dynamic sectors like network industries. It seems an impossible task to design a complete regulatory contract, without having to modify it frequently. Not only will legal adjustment be slow and costly, but it is also likely to open up opportunities for strategic behaviour by the regulated firms. Moreover, if the law contains too much detail the regulator will be vulnerable to legal challenge. On the other hand, if the law is too broad (degree of detail too low) it is likely to be incomplete. This implies that either the regulation may take an unintended shape or the regulator should be given the authority to interpret the law: i.e. the regulator should have flexible powers. Hence, independence requires flexible powers.

The story of the regulated (electricity) charges in the Netherlands illustrates well.[31] The regulator for energy markets in the Netherlands is DTe which is a chamber of the competition agency NMa, which is a supervisory body of the Ministry of Economics. DTe decided correctly to apply individual X_i's to firms for the first regulation period (2001 - 2003), after which it could reasonably be expected that firms would have caught up and the non-individual X yardstick could then be applied in the second regulation period. The X_i's determined by DTe in the first round were high. The sector did not accept them and appealed. Importantly, appeals against the DTe decisions are the responsibility of the Court of Appeal for Trade and Industry (CBB), which makes judicial reviews rather than a testing for

[30] The terms are derived in analogy to the management-shareholder relation as set out in Vickers & Yarrow (1988, p. 21) who use the term 'managerial discount factor'.

substance. It decided that the law did not allow individual X-factors and the regulation had to be revised (retrospectively). This illustrates the flaw of the system very well. On the one hand, one would like to give DTe flexible powers to interpret the law, but on the other hand one would like to constrain DTe's authority. Since there is no substance review, the judicial review by the Appeal Court restricted DTe's power to interpret the law, at the expense of good substance.[32]

Two approaches mitigate the apparent incompatibility of independence and flexible powers. First, a system of checks and balances. Second, changing the incentives of the regulator. A system of checks and balances has been developed in, for example, the UK (cf. Green, 1999 and Geroski, 2004),[33] between especially the regulators and the competition commission (CC). This sounds easier than it actually is. Green (1999) points out that regulators have to follow CC's recommendations following a referral. The inevitable result is that the regulators start to anticipate and mimic the CC's procedures, implying that the system of checks and balances is undermined. A strong system of checks and balances requires different institutions having different mandates and possibly accentuating different interests (cf. for a similar argument, Laffont & Martimort, 1999). This, however, runs into the problem of who decides in case of disagreement. The following strikes a balance. First, regulators should have a clear mandate, phrased preferably in terms of meaningful and possibly well-defined objectives. A court can (should) then test whether the regulator acts according to the objectives set by the legislator. Second, a system of checks and balances of different institutions, controlling each other with non-binding recommendations, would guarantee a check on substance.

The relation in the UK between the CC and the sector regulators illustrates well. Geroksi (2004) explains the two-tier approach with the sector regulators in the first tier and the CC in the second tier. According to Geroski (2004, p. 78, italics added), the CC is: "an investigative body with the expertise to decide on issues of *substance*." Interestingly, the CC has determinative powers for most regulators, but in the case of the regulation airport landing charges it can only make non-binding recommendations. Where CC decisions are determinative, firms can appeal for judicial review by the Competition Appeal Tribunal.

Alternatively, the incentives of the regulator may be recognised and if necessary changed. In a way, the approach of increasing independence is an 'incentive mechanism'; the

[31] The interested reader may be referred to Hancher, et. al. (2003) and Nillesen & Pollitt (2004) for more detail.
[32] The ruling had several consequences. First, the law was modified retrospectively to allow individual X-factors. Second, the X-factors had to be adjusted again. Finally in May 2003, the final X factors were agreed with the sector; this was only months before the end of the regulatory period.

idea is to reduce the incentives to behave opportunistically (or myopically) by lengthening the 'regulatory discount factor'. Further, the analysis of UUROR by Gilbert & Newbery (1994) as set out above, can be seen as an incentives change. They explain that allowing the clause that assets should be 'used and useful' reduces the incentives of the regulator to deviate from the (optimal) regulatory contract after the investment. Lastly, suppose that complete regulatory independence from political interference is an illusion and suppose further that it is correct that the state's incentives to capture the regulator depend on ownership of the regulated industry. If the state is owner, the state will want to have high prices, while if the industry is privately owned the state will press for low prices. The logical conclusion must be that private-public ownership might strike a balance on the incentives of the state. This view adds another dimension to the privatisation debate.

4 Concluding remarks

This paper identifies and analyses key sources of policy uncertainty which impede investment incentives in the electric power industry. New investment critically determines long term system reliability, which is captured by the term supply adequacy, covering both generation and network adequacy. The benchmark in this paper is that symmetric, non-systematic, diversifiable risk without (future) spin-offs has no effect on investment. We have constrained attention to 1) asymmetric uncertainty, 2) non-diversifiable risk and 3) uncertainties creating a real option value. We use the term policy uncertainty in a broad sense to cover the effects of uncertain regulation where it already exists, the effects of market risks under different types of regulation and, importantly, the mere possibility of an intervention where no policy exists or where policy is vague.

The evidence on generation adequacy is inconclusive. The short-run outlook is optimistic as are stress-tests, but the long-run perspective (although not entirely reliable) gives reason for concern. Calculations suggest that hands-off, energy-only markets and resulting prices might simply not suffice to attract sufficient new investment in peaking plant. That prices would have to be implausibly high is supported by three observations. Firstly, it has not yet happened that prices have been high enough and for sufficient duration to cover estimated costs of peaking generation capacity. Second, as discussed by Joskow and Tirole (2004) high prices are extremely sensitive to the discretion of the TSOs. Third, extreme situations are likely to trigger intervention. Hence, the mere possibility of an intervention already

[33] See Vol. 12 (2004) of Utilities Policy which is a special issue on the UK.

asymmetrically caps revenues. Following this line of argument, it seems that the policy expectations can be self-fulfilling leading to multiple potential equilibria: low capacity equilibria with interference and high capacity equilibria without interference.

If the scenario of fulfilled policy expectations is plausible, the second-best option of actually doing something seems inevitable. Hence, a 'better safe than sorry' policy may simply be realistic. To this end, a myriad of capacity mechanisms have been proposed and debated in the literature. Part of this debate centres around the relative merits of price and quantity methods of incentivizing investment. As discussed by Stoft (2002) capacity payments produce higher, more erratic price spikes implying that quantity measures have less inherent risk and are less susceptible to market power abuses. As such one would also expect quantity measures to invite less regulatory interference. Capacity mechanisms should be flexible and 'market-friendly' including well-defined long term and secondary contract markets. Reserve pricing fits this description and can include hedging instruments that might be exercised centrally by the system operator or bilaterally between producers and load. A disadvantage relative to capacity 'obligations' is that the SO has discretion over both the quantity of reserves and the willingness to pay for reserves. With obligations only the quantity is determined by the SO. Future policy should address ways to lower regulatory discretion and minimize distortions from capacity markets on energy market prices.

The situation for network adequacy is rather different than for generation. With the exceptions of the interconnectors, network investment seems to be good empirically. Theoretical concern over underinvestment in networks may be empirically negligible or not yet observable. With respect to policy uncertainty there are two lines of argument. First, price cap regulation is likely to increase market risk compared to no regulation and especially rate of return regulation. Second, regulators may be unable to credibly commit to a consistent regulatory policy and to refrain from opportunistic behaviour. The 'hold-up' or 'non-commitment' problem has received extensive attention in the literature. This is particularly relevant for significant new investment under price caps. Importantly, Gilbert & Newbery (1994) argue that the so-called used-and-useful rate-of-return regulation (UUROR) may be seen as a way to increase the regulator's credibility and would therefore improve investment incentives. Overall, we would expect to see rate-of-return regulation regaining territory compared to price capping if underinvestment in the network gains importance and short-run efficiency gains are exhausted; of course, shifts are likely to find subtle channels like smaller regulatory lags and profit-sharing clauses.

Lastly, we discussed the institutional prospects to improve the credibility of the regulators in order to improve the investment climate. Given the key importance of independence from political interference and flexible powers, a system of checks and balances with an institution with jurisdiction on substance is recommendable. Although debatable, it may well be that such an agency should have recommendation powers only.

References

Armstrong, M., Cowan, S. & Vickers, J., 1994, *Regulatory reform; Economic analysis and British experience*, MIT Press, Cambridge, Mass..

Bialek, J.W., 2004, 'Recent blackouts in US and Continental Europe: Is liberalisation to blame?' *Working Paper* CMI EP 34, University of Cambridge.

Bijvoet, C.C., Nooij de, M. & Koopmans, C.C., 2003, 'Kosten van stroomstoringen', *ESB*, 3-10-2003, pp. 460-463.

Brattle Group, 2004, *Long-term reserve contracts in the Netherlands*, Brattle Group, London.

Brealey, R.A. & Myers, S.C., 2003, 7th. ed., *Principles of corporate finances*, McGraw Hill.

Brennan, M.J. & Schwartz, E.S., 1982, 'Consistent regulatory policy under uncertainty', *Bell Journal of Economcs*, Vol. 13, No. 2, pp. 506-521.

Brunekreeft, G., & Twelemann, S., 2004, 'Regulation, competition and investment in the German electricity market: RegTP or REGTP', *Working Paper* CMI EP 68, University of Cambridge, forthcoming *Energy Journal*.

Brunekreeft, G., 2004, 'Regulatory issues in merchant transmission investment:', Working Paper CMI-EP-38, University of Cambridge, forthcoming *Utilities Policy*.

Brunekreeft, G., Neuhoff, K. & Newbery, D.N., 2004, 'Electricity transmission: An overview of the current debate', *Working Paper* CMI EP 60, University of Cambridge.

Bushnell and Mansur, 2004, "Consumption under noisy price signals: a study of electricity retail rate deregulation in San Diego," *Power Working Paper-082r*, University of California Energy Institute, Berkeley.

Cabral, L.M.B., 2000, *Introduction to Industrial Organization*, MIT Press, Cambridge, Mass..

CEC, 2002, 'Second benchmarking report on the implementation of the internal electricity and gas market,' *Commission Staff Working Paper*, Oct. 2002, SEC(2002)1038, European Commission, Brussels.

Clingendael Institute, 2004, *Study on the energy supply security and geopolitics*, Report prepared for DG TREN, January 2004, The Hague.

CPB, 2004, 'Better safe than sorry? Reliability policy in network industries,' *CPB document* No. 73, The Hague.

Dixit, A.K. & Pindyck, R.S., 1994, *Investment under uncertainty*, Princeton University Press.

Fehr von der, N.-H., Amundsen, E.S. & Bergman, L., 2004, 'The Nordic market: Signs of stress?, forthcoming *Energy Journal*, *mimeo*, July 12, 2004.

Gans, J. & King, S., 2003, 'Access holidays: The panacea for network infrastructure investment?', *mimeo*, March 2003, University of Melbourne.

Geroski, P.A., 2004, 'Appealing to the UK competition commission', *Utilities Policy*, Vol. 12, pp. 77-81.

Gilbert, R.J. & Newbery, D.M., 1994, 'The dynamic efficiency of regulatory constitutions', *Rand Journal of Economics*, Vol. 25, No. 4, pp. 538-554.

Green, R., 1999, 'Checks and balances in utility regulation - The UK experience', *Public Policy for the Private Sector*, World Bank Series No. 185, May 1999, World Bank.

Grout, P.A. & Zalewska, A., 2003, 'Do regulatory changes affect market risk?', LIFE Working Paper 03-022, LIFE, Maastricht University.

Hancher, L., Larouche, P. & Lavrijssen, S., 2003, 'Principles of good market governance', *Journal of Network Industries*, Vol. 4, pp. 355-389.

Helm, D., Hepburn, C. & Mash, R., 2003, 'Credible carbon policy', *Oxford Review of Economic Policy*, Vol. 19. No. 3, pp. 438-450.paper on CO2 OXREP

Hirst, E., 2004, *U.S. transmission capacity: present status and future prospects*, June 2004, Washington DC.

JESS, 2003, *Joint energy security of supply working group (JESS); Third report*, November 2003, London.

Joskow, P.L. & Tirole, J., 2004, 'Reliability and competitive electricity markets', *Working Paper*, CMI EP 53, University of Cambridge.

Joskow, P.L., 1974, 'Inflation and environmental concern: Structural change in the process of public utility price regulation', *Journal of Law and Economics*, Vol. 17, No. 1, pp. 291-327.

Joskow, P.L., 1989, 'Regulatory failure, regulatory reform, and structural change in the electric power industry', *Brookings Papers on Economic Activity; Microeconomics*, Vol. 1989, pp. 125-208.

Joskow, P.L., 2003, 'The difficult transition to competitive electricity markets in the U.S.,' MIT, May 2003.

Laffont, J.-J. & Martimort, D., 1999, 'Separation of regulators against collusive behavior', *Rand Journal of Economics*, Vol. 30, No. 2, pp. 232-262.

Lyon, T.P. & Mayo, J.W., 2000, 'Regulatory opportunism and investment behavior: evidence from the US electric utility industry', *mimeo*, June 2000.

Myers, S.C., 1972, 'The application of finance theory to public utility rate cases', *Bell Journal of Economics*, Vol. 3, No. 1, pp. 58-97.

Neuhoff, K. & De Vries, L., 2004, 'Insufficient incentives for investment in electricity generations', *Utilities Policy*, Vol. 12, pp. 253-267.

New England ISO, Web page: http://www.iso-ne.com/

New York ISO, Web page: http://www.nyiso.com/

Nillesen, P.H.L. & Pollitt, M.G., 2004, 'The consequences for consumer welfare of the 2001-2003 electricity distribution price review in the Netherlands', *Working Paper*, CMI EP 50, University of Cambridge.

Ocana, C., 2003, 'Trends in the management of regulation: a comparison of energy regulators in member countries of the OECD', *International Journal of Regulation and Governance*, Vol. 3, No. 1, pp. 13-32.

OECD, 2002, *Security of supply in electricity markets*, OECD, Paris.

Oren, S., 2004, 'Ensuring generation adequacy in competitive electricity markets,' *mimeo*, UC Berkeley, April 2004.

Peltzman, S., 1976, 'Toward a more general theory of regulation', *Journal of Law and Economics*, Vol. 19, No. 2, pp. 211-240.

PJM ISO, Web page: www.pjm.com

PJM Market Monitoring Unit,. Web page: Web page: http://www.pjm.com/markets/market-monitor/market-monitor.html

Roques, F., Newbery, D.M. & Nuttall, W.J., 2004, 'Generation adequancy and investment incentives in the UK: from the pool to NETA', *Working Paper*, CMI EP 58, University of Cambridge.

Spence, A.M., 1975, "Monopoly, quality and regulation", *Bell Journal of Economics*, pp. 417-429.

Stoft, S., 2002, *Power system economics*, Wiley Interscience.

TenneT, 2004, *Rapport Monitoring Leveringszekerheid* 2003-2011, June 3, 2004, TenneT, Arnhem.

U.S Department of Energy, Energy Information Association, www.eia.doe.gov.

UCTE, 2003, *UCTE system adequacy forecast 2004-2010*, report, December 2003, UCTE.
UCTE, 2004, *UCTE system adequacy retrospect 2003*, report, June 2004, UCTE.
Vázquez, C., Rivier, M. & Pérez-Arriaga, I.J., 2002, 'A market approach to long-term security of supply', *IEEE Transactions on Power Systems*, Vol. 17, No. 2, pp. 349-357.
Vickers J. & Yarrow, G., 1988, *Privatization: An economic analysis*, MIT Press, Cambridge, Mass..
Vries de, L., 2004, *Securing the public interest in electricity generation markets: The myths of the invisible hand and the copper plate*, TU Delft, Netherlands.
Wright, S., Mason, R. & Miles, D., 2003, *A study into certain aspects of the cost of capital for regulated utilities in the U.K.*, Report, February 13, 2003, Smithers & Co, London.