24 June 2005

Investment Incentives for the SE Europe Regional Electricity Market – Volume 1

The World Bank

NERA
Economic Consulting

MMC Marsh & McLennan Companies

Project Team

Graham Shuttleworth
Sean Gammons
Hamish Fraser

NERA Economic Consulting
15 Stratford Place
London W1C 1BE
United Kingdom
Tel: +44 20 7659 8500
Fax: +44 20 7659 8501
www.nera.com

Contents

EXECUTIVE SUMMARY

South Eastern Europe (SEE) is moving toward integration of electricity markets. According to most studies the sector will need investment in new generation capacity over the next five to ten years. To encourage this investment, SEE governments need to consider providing clear incentives that will encourage private sector investors to commit funds to long-term investment in the region. At the moment, many markets in the SEE region are unlikely to provide these incentives.

In future, it will also be necessary to find some way to co-ordinate investment in the region as a whole, principally to ensure that one country does not, either deliberately or accidentally, rely too heavily on the generation capacity of its neighbours.

We have reviewed a number of different ways of encouraging investment in capacity, specifically use of Power Purchase Agreements (PPAs) issued by a single buyer, an energy-only market and various forms of capacity payment or obligation. We have reviewed these schemes according to the following criteria:

1. Security of supply (adequate level of investment)
2. Cost efficiency:
   - appropriate risks to investors (credit risk, price risk, regulatory risk)
   - efficient long-term decisions on investment (how much capacity to build/ retire, what type of capacity and where and when to do it)
   - Efficient short-term price signals (particularly, avoid distortions to dispatch)
3. Competition - ability of traders and consumers to choose who supplies them and on what terms
4. Transparency – not necessarily simplicity, but predictability and stability of rules
5. Price stability to customers
6. Non-susceptibility to gaming
7. Fairness

Overall we have found that the use of PPAs scored highly according to these criteria. However, the single buyer approach is inconsistent with promotion of competition as per the latest EU Directives on the internal electricity market, and with the stated long-term aims of the Athens Forum.

Energy-only markets are consistent with European requirements for competition but we have found that they provide an unreliable form of investment incentive especially in the conditions found in the SEE region. The incentive for investment in these markets comes from price spikes – very high prices that occur when demand exceeds capacity. These price spikes happen only occasionally and cause problems for consumers and regulators when they do. On the other hand they may be absent for many years in which case investors have trouble recovering their costs. To overcome this kind of “boom-and-bust” cycle, energy-only markets work best in conjunction with long-term contracts which stabilise charges to consumers and revenues to investors. The task facing any electricity market designer is to ensure that retailers provide the necessary long-term commitments to investors by signing long-term contracts or offering some equivalent alternative.

We have investigated a number of centrally coordinated capacity schemes: (1) capacity payments or "adders" which augment revenues from sales of energy; (2) capacity obligations which require retailers to sign contract or incur a penalty; and (3) a hybrid scheme which makes the volume of the obligation vary in response to the observed price of capacity.

In the conditions of SEE, we found it would be difficult to set a capacity payment that ensured construction of the right level of capacity, since the cost of building capacity is subject to wide variations associated with local costs and country risk. A small error in estimating the cost could result in large over- or under-investment.

A capacity obligation defined in quantity terms reduces the scope for errors in that regard. However, a fixed capacity obligation is only as effective as the penalty used to enforce it and such schemes have produced high and volatile prices for capacity in other countries. To minimise such price fluctuations, we concluded that capacity obligations, where they are adopted, must be monitored and enforced in advance, because forward prices for electricity fluctuate less than spot prices. Thus, we recommend that retailers should be required to show that they have secured capacity at least one year in advance. Based on recent innovations in US capacity markets, we also recommend that the penalty for falling short of the obligation should be defined by a price-responsive demand for capacity. Both these features would help to dampen price fluctuations and would cause prices to rise gradually as the market tightens, instead of swinging from a zero value for capacity to a penalty value and back again.

However, for any of these schemes to work investors must first be confident that their customers and the industry as a whole will be creditworthy. Hence, the priority is to ensure that tariffs cover costs. Without meeting this pre-condition, it will be difficult to encourage any private sector investment in SEE member countries.

Assuming that this pre-condition can be met, the members of the SEE regional electricity market need to work towards a harmonised (though not necessary common) system of capacity obligations. Moving towards a harmonised system requires the following steps.

1. Reach a regional agreement to allocate capacity obligations to each SEE country
2. Within each SEE country, allocate the national obligation among one or more parties (TSOs or retailers), if necessary by zone (where there are significant transmission constraints)
3. Set up a scheme in each zone for measurement and registration of capacity holdings (including bilateral arrangements for accepting imports and exports of capacity)
4. Define the terms of the capacity contract that is accepted against the obligation (i.e. what generators have to do)
5. Develop a method for TSOs to buy capacity to make up shortfalls of individual retailers (and to charge the retailers for it)
6. Ensure that the costs are passed through to customers

Of these tasks, 1 and 6 have a priority, since they will determine whether the investment climate for region as a whole will ensure security of supply within the coming years.

1. Introduction

This study has been commissioned by the World Bank and is intended to assist the member states of the South East Europe Regional Electricity Market to attract investment into their electricity systems.

The region of South East Europe (SEE) is characterized by a number of electricity markets in various stages of early development. The market pricing mechanisms adopted in these countries have had a mixed record in encouraging long-term investment in new electricity generation capacity. The retail market is also facing a number of reform challenges, particularly a need to raise prices up to the level of costs, and to provide the conditions for consumers to choose their suppliers.

These problems are not all unique to South East Europe. Electric markets in other parts of the world have already been restructured, but are still struggling to find a method of ensuring supply adequacy and protection against undue price volatility. The sources of this concern vary throughout the world, and include retail prices set by governments far below the cost of new entry, and the need to compensate for politically imposed price caps in energy markets. A key concern in the context of SEE is that a market for electrical energy – by itself – might not provide for sufficient assurance of security of supply.

This report investigates alternative investment support mechanisms, called capacity schemes, which might provide the necessary support to the energy market in the South East Europe context. This report provides an overview of conditions in the electricity markets of South East Europe and then considers the different ways in which electricity markets encourage investment in capacity. Chapter 3 sets out the basic economics of capacity incentives, whilst chapter 4 considers the merits of particular schemes. The analysis of different schemes suggests that energy-only markets will not meet the needs of the region, primarily because they rely on occasional price spikes to provide incentives for investment. Such occasional price spikes are likely to be unsustainable in the region and unattractive to investors as the basis for long-term commitments. Chapter 6 concludes with a recommended path from establishing basic preconditions, through central procurement of capacity via PPAs, to the outline of a capacity obligations scheme.

2. The SE Europe Regional Electricity Market

This section provides an overview of the electricity sector in the countries that make up the South East Europe (SEE) regional energy market.[1] Sections 2.1 to 2.5 describe the current market with reference to the level of competition for supply, the outlook for investment plans, industry structure, fuel choices and imports and exports. Sections 2.6 and 2.7 focus on the forecast state of the market both in terms of capacity growth and growth in transmission capabilities.

2.1. State of the Market

Table 2.1 shows electricity consumption, peak demand, installed capacity, and net imports in the relevant markets except for Turkey, and UNMIK.

Table 2.1
Energy Market Indicators 2000[2]

	Electricity Consumption GWh	Peak Power Demand GW	Installed Capacity GW	Net imports +/- GWh	Import Penetration %
Albania	3,639	7.50	1.6	1,002	28%
Bosnia and Herzegovina	5,860	1.00	3.9	-1,064	-18%
Bulgaria	25,486	5.10	12.6	-4,619	-18%
Croatia	12,206	1.90	3.9	3,999	33%
FYR Macedonia	5,408	1.00	1.7	112	2%
Greece	44,974	8.70	11.8	-11	0%
Romania	41,293	7.50	23	-696	-2%
Serbia and Montenegro	26,703	4.80	12.6	3,908	15%
Total	165,569	38.00	68	2,631	-

The table shows that import penetration and therefore interconnection is relatively high in some countries i.e. Croatia and Albania but relatively low in others i.e. Macedonia and Romania. The total capacity of interconnectors in the Eastern European Area (EEA)[3] is estimated at 6 GW whilst the total capacity of interconnectors between the main area and the periphery is estimated at 12 GW.[4,5] Total capacity for the area is 68 GW[6] indicating that overall there is a relatively low level of interconnectivity both between the main area and periphery.

---

[1] The SEE includes the Republic of Albania, Bosnia and Herzegovina, Republic of Bulgaria, Republic of Croatia, Republic of Greece, Republic of Romania, Republic of Turkey, the State Union of Serbia and Montenegro, Former Yugoslav Republic of Macedonia and the United Nations Interim Administration Mission in Kosovo (UNMIK).

[2] Professor Pantelis CAPROS, Towards a Regional Energy Market for the South-East European Area, National Technical University of Athens, April 2005, page 3.

[3] The EEA includes Romania, Bulgaria, Serbia & Montenegro, Kosovo, Bosnia-Herzegovina, Croatia, FYR Macedonia, Albania, and Greece.

[4] The periphery includes Turkey, Italy, Austria, Slovenia, and Hungary.

[5] Professor Pantelis CAPROS, Towards a Regional Energy Market for the South-East European Area, National Technical University of Athens, April2005, page 4.

[6] See Table 1.1.

2.2. Industry Structure

Industry structure is developing in all SEE markets at varying speeds. According to a set of benchmarks laid out in the Athens Memorandum and assessed by the World Bank Romania has made the most reform progress followed by Turkey, Bulgaria, and Croatia.[7] The relevant benchmarks are: effective tariffs and affordability; regulation; and industry restructuring.

The following indicators based the World Bank Framework for Development of Regional Energy Trade in South East Europe[8] illustrates the development of industry structure in the region:

§ Independent regulators are in place in Albania, Bulgaria, Croatia, Romania and Turkey. Legislation to enable the set up of independent regulators has been adopted in Bosnia, Macedonia, and Montenegro. Serbia has drafted the relevant legislation.

§ Cross subsidies between industrial and residential customers is an issue in Albania and residential to industrial tariff ratios are relatively low in Bosnia, Bulgaria, Croatia, Romania and Turkey.

§ Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Kosovo, Romania, and Serbia all have in place a social safety net which is typically in the form of a block tariff for basic consumption.

§ Bulgaria, Romania, and Turkey have adopted transmission and distribution tariff methodologies.

§ In Bulgaria and Romania vertical unbundling has taken place, i.e. the separation of generation and transmission companies. In Bulgaria horizontal unbundling has also taken place with the separation of some thermal generation.

§ There has been some private sector involvement in the region including the sale of a number of distribution companies in Romania. However, private sector interest has been diminished due to poor payments discipline in the area. Payments discipline is a significant problem in most SEEREM countries and the average collection ratio for the region is 85%.

Wholesale market arrangements vary widely between SEE countries. Greece is about to adopt a mandatory day-ahead pool, but other countries in the region are operating on the basis of bilateral contracts and a variety of balancing mechanisms for handling flows outside contracts.

2.3. Competition for Supply (Retail)

The European Directive on the internal market in electricity requires that members of the European Union open up electricity markets to competition for supply to all non-household

---

[7] David Kennedy and John Bessant-Jones, World Bank Framework for Development of Regional Energy Trade in South East Europe, Energy and Mining Sector Board Discussion Paper, Paper no.12, March 2004, Pages 15, 29 and 31.

[8] Kennedy and Bessant-Jones (2004), Pages15, 29 and 31.

customers by July 2004 and to all customers by July 2007.[9] The only SEE country which is a member of the European Union at present and thus subject to these conditions is Greece. However, Bulgaria and Romania are expected to become members of the EU in 2007. For other SEE countries, the latest documents on the regional electricity market envisage that retail competition in the remaining states will be extended to non-households by 2008 and to all households by 2015.[10]

A major problem for the electricity sector is the low level of regulated tariffs in many markets. Low tariffs not only prevent any competitor from entering the business, they also mean that some of the existing providers are failing to cover their costs. In these conditions, such companies do not make a creditworthy partner for new contracts, but there is political resistance to major increases in retail electricity prices.

2.4. Fuel Choices

In the EEA over half of electricity generation was generated from thermal energy in 2002. The remainder was generated from nuclear and hydro, which accounted for approximately 25% and 20% of total generation respectively.[11] Coal is an important thermal fuel source for power generation and is expected to continue to be, although increasingly gas may replace coal in new generating plants.[12] The largest coal industries in the region are found in Bulgaria, Serbia, and Romania, which each produce around 30 million tonnes per year. Kosovo and Bosnia and Herzegovina also have coal industries, each producing around 10 million tonnes per year. Kosovo holds the largest reserves of low cost lignite in the region.

The level of gasification in SEE is relatively low with combined demand from Albania, Bosnia and Herzegovina, Macedonia, and Serbia totalling less than 1 bcm per year.[13] Montenegro and UNMIK have no gas sector at all. There is some indigenous gas production in the region. Romania has the most significant domestic production; Croatia and Serbia also have reserves although Croatia's are declining.

2.5. Imports and Exports

Table 2.1 shows the electricity trading positions of SEE member states. Bulgaria, Bosnia and Herzegovina, and Romania are all net exporters of power. Bulgaria needs to maintain high reserves to cover its large nuclear units. As a result, the Bulgarian electricity system has many opportunities to export to neighbouring countries and could also share reserve with its

---

9 European 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.

10 European Commission, DG TREN, South East Europe Electricity Market options paper, draft v.1, Brussels, 12 April 2005.

11 Energy Information Administration, Country Analysis Brief: South Eastern Europe, March 2004, Page 5.

12 David Kennedy and John Bessant-Jones (2004), Page 21.

13 David Kennedy and John Bessant-Jones (2004), Page 18.

neighbours, which would lower total costs.[14] Bosnia and Herzegovina currently has a reserve margin of 97% and will probably continue exporting at peak times to the regional market.[15]

Albania has a large proportion of hydro generation, which means it has potential to export power in wet years but must import in dry years. However, limited transmission capacity is currently limiting its ability to trade. Imports to Macedonia are expected to increase as domestic coal reserves are exhausted.[16] Turkey, which is currently a net importer, could become a net exporter over the medium term as technical constraints to trade diminish.[17]

In some cases, there is additional potential for trade, but a number of problems prevent trade from happening, including poor information about these opportunities, limited transmission capacity, and possibly credit risk. Fear of non-payment by existing utilities in neighbouring countries may also be harming investment in generation and transmission capacity.

2.6. Forecast State of the Market

Table 2.2 shows a recently published forecast of demand and electricity capacity for participating markets (except Turkey) based on three scenarios:[18]

§ Scenario A provides a least-cost plan for investment, in the extreme case where each national power system builds enough capacity to meet its own demand independently, taking into account the opportunities for trade over existing interconnectors.[19]

§ Scenario B describes least-cost investment in a fully integrated regional system without any physical or trade constraints.

§ Scenario C is an intermediate case including the impact of proposed regional interconnector capacity on generation capacity.

The table shows that between 2005 and 2020, capacity requirements increase by 26% under scenario A and by 14%–15% under scenario B. Under scenario C, capacity increases by anything from 3% to 31%, depending on assumptions about demand, gas prices, increased hydro generation, rehabilitation of existing plants, growth in nuclear capacity, increased construction costs and discount rates.

Data for Greece is not contained in the same source. However, a recent study of requirements shows that Greece is expected to remain a net importer over the period 2010–2015.[20]

---

14 David Kennedy and John Bessant-Jones (2004), Page 10.

15 David Kennedy and John Bessant-Jones (2004), Page 9.

16 David Kennedy and John Bessant-Jones (2004), Page 9.

17 David Kennedy and John Bessant-Jones (2004), Page 10.

18 The European Union's CARDS programme for the Balkan region – Contract No. 52276, REBIS: GIS, Draft Final Report, 31 December 2004.

19 For Bulgaria and Romania, these plans take into consideration full environmental compliance with the Large Combustion Plants Directive (Directive 2001/80/EC of the European Parliament and of the Council of 23 October 2001 on the limitation of emissions of certain pollutants into the air from large combustion plants). For other countries, the forecasts envisage limited environmental compliance.

Table 2.2
Forecast Electricity Demand, Capacity and Peak Load[21]

	Compound annual growth in gross electricity demand	Forecast total system net capacity		
	%	MW	MW	MW
	2003 - 2020	2005	2010	2020
Regional Total	2.3			
Regional Total - Scenario A		42,277	44,224	53,158
Regional Total - Scenario B		42,817	41,027 - 41,423	48,625 - 49,128
Regional Total - Scenario C		42,654 - 42,817	40,653 - 42,935	44,304 - 55,762

I

nvestment plans vary across the region under each Scenario. Table 2.3 illustrates capacity in 2002 and forecast of capacity investment in each country in the region according to Scenario A. In this Scenario, Romania, Bulgaria, and Albania are already committed to construction of new generation units, but other countries also require investment in due course.[22]

Table 2.3:
Forecast Capacity Across the SEE Region

	Forecast System Net Capacity (GW)			Forecast Capacity Increase (%)
	2005	2010	2020	2005 - 2020
Bulgaria	9.38	9.40	9.44	1%
Serbia (excluding UNMIK)	7.20	7.83	8.31	15%
Romania	14.04	13.42	16.57	18%
Bosnia and Herzegovina	3.49	3.54	4.22	21%
Croatia	3.75	4.27	5.03	34%
FYR Macedonia	1.46	1.70	2.79	91%
UNMIK	1.07	0.91	2.12	99%
Albania	1.45	2.30	3.26	125%
Montenegro	0.45	0.86	1.41	213%
SEE Total	**42.28**	**44.22**	**53.16**	**26%**

Source: The European Union's CARDS programme for the Balkan region – Contract No. 52276, REBIS: GIS, Draft Final Report, 31 December 2004, section 3a.

[20] Professor Pantelis CAPROS, Towards a Regional Energy Market for the South-East European Area, National Technical University of Athens, April 2005, page 8.

[21] The European Union's CARDS programme for the Balkan region – Contract No. 52276, REBIS: GIS, Draft Final Report, 31st December 2004. Volume 1; table 1, volume 3a; tables 32 and 36 – 43.

[22] The European Union's CARDS programme for the Balkan region – Contract No. 52276, REBIS: GIS, Draft Final Report, 31 December 2004, section 3a.

2.7. Expected Increases in Transmission Capacity

A number of SEE states are planning to increase interconnection capacity. Participants in the Athens Memorandum have proposed a number of reconnection, strengthening, and completion projects for interconnectors in the region. The projects specifically identified by the Athens Forum Process are;

- the reconnection of lines in Croatia /Serbia / Bosnia and Herzegovina;
- the completion of the Western North-South Connection through Elbasan in Albania;
- the strengthening of the Greece-Bulgaria links; and
- the strengthening of the Greece-Italy and the Greece-Turkey links.

Completion of these projects will be contingent on ever increasing political co-operation between various states within the SEE region.

2.8. Conclusions

The energy markets in the SEE vary according to a number of indicators including capacity and interconnection, competition for supply, industry structure, fuel choices, and import and export penetration.

Demand in the region is expected to increase at a compound rate of 2.3% per annum over the period 2003–2020. Capacity is required to grow at the same time and particularly over the period 2010–2020 although this growth in capacity will not be evenly spread across the region. Furthermore, the required volume and cost of this expansion is unclear. It will depend on the degree to which the markets are interconnected, with increased integration requiring smaller increases in capacity. However, the degree of integration will depend upon the willingness of SEE countries to trade with one another in the short and the long term - i.e. to trade electricity and to rely on each other's capacity.

3. Conceptual Basis for Capacity Schemes

Appendix A sets out a market paradigm which shows how energy markets could, in principle, provide a reward for investing in capacity, by paying very high prices during times of shortage. Volume 2 of this report shows that at least some electricity markets are currently functioning without any explicit scheme for remuneration of generation capacity. It is therefore important to understand why electricity markets might deviate from the market paradigm, or why the outcome of energy-only markets might be viewed as inadequate. Only then is it possible to understand (1) the reasons for setting up explicit capacity schemes and (2) what design features meet these needs.

3.1. The Market Paradigm

The basis for "energy-only" electricity markets is the assumption that product prices will encourage investment in generation capacity, just like prices do in markets for other products.

Box 1: Illustrative Example of "Energy-Only" Market Prices

The Table below provides an illustrative example of the effect that capacity shortages have on average electricity market prices. The first row in the table shows the number of years out of 10 in which a shortage occurs (1 year out of 10 on the left, rising to 10 years out of 10 – i.e. every year – on the right). The second row shows how many times the demand should reach the available capacity in these shortage years to achieve a long-run average of 10 hours per year. If only one shortage year occurs every 10 years, that year must contain all 100 hours of the prices equal to VOLL needed for the whole 10 year period. If there are two shortage years, each must have 50 hours of the maximum prices.

Assume that the annual fixed cost of a peaking generator is €50/kW per year and that the desired total duration of maximum price periods is 10 hours per year on average. These figures are consistent with a Value of Lost Load of €5,000/MWh (a typical value). Suppose that the average price in other hours is €20/MWh. The bottom line of the table calculates the average electricity price that must apply in shortage years, depending upon the number of shortage years on average in each 10 year period. It shows that spreading that reward for capacity over all 10 years would increase the average electricity price to €26/MWh. However, if the reward for capacity arises in only one year out of 10, prices in that year must rise to €77/MWh on average.

Average Electricity Prices in Energy-Only Markets (Example)

Number of Shortage Years out of 10		1	2	3	4	5	6	7	8	9	10
Number of Hours with the Maximum Price	Hours	100	50	33	25	20	17	14	13	11	10
Maximum Electricity Price	€/MWh	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
Average Price in All Other Hours	€/MWh	20	20	20	20	20	20	20	20	20	20
Average Electricity Price in Shortage Years	€/MWh	77	48	39	34	31	29	28	27	26	26

As Appendix A explains, this market paradigm assumes that prices rise to very high levels when there is a shortage, in order to ration scarce supplies of electricity. When demand threatens to exceed available capacity, the price must rise to the "Value of Lost Load" (VOLL) (also known as the "Cost of Unserved Energy"), for which estimates lie anywhere in the range €2,000-€10,000/MWh - almost 1,000 times the price of electricity in "normal" conditions. *On average*, the price of electricity must reach this level for several hours per year to encourage investment in generation capacity. In practice, such high price episodes may be concentrated in one or two years, (for instance, in the years of the extreme temperature conditions, outages of major generation units, or failures of major transmission lines) and be almost absent in other years.

Driving electricity prices up to VOLL when a shortage occurs is one way to remunerate investment. Some markets operate on this basis and have seen extreme fluctuations in market prices as a result. (In hydro systems, the price fluctuations are not so extreme, but last longer, so that the required average prices are the same.) However, it is known that many markets are subject to constraints which prevent prices from rewarding investment in this way.

3.2. Constraints on the Market Paradigm

3.2.1. Market design features

Many documents mention how "electricity cannot be stored", and "production must balance demand at all times", and "investment in generation is highly risky", but these "special conditions" apply in other markets and/or provide no clear reason for giving special treatment to investment incentives in electricity markets. However, electricity markets do suffer from two important conditions which affect market design:

1. It is possible for consumers (or their suppliers) to take a supply of electricity from the network, even if they have no contract with a generator (or if their demand exceeds their contracted capacity); and
2. Outages during capacity shortages cannot be targeted on those customers whose demand exceeds their contracted capacity.

The first of these conditions leads to creation of a "settlement system" - a mandatory pool or obligatory "balancing mechanism".[23] These settlement systems allow the Transmission System Operator (TSO) to arrange short-term production of electricity to ensure that production always balances demand, and charges market participants for any electricity they take outside their contracts (or pays them for any surplus production). The pricing rules in these settlement systems have to be determined centrally by a formula and are often subject to regulation (since the settlement system is a monopoly).

---

[23] A mandatory pool is a market that settles all physical production, or all physical injections into a network. In such a market, contracts have to be financial "contracts for differences", although some pools allow generators to set the level of their output entering the pool equal to the volume of contract sales. A balancing mechanism allows market participants to net off the volume of their contract sales and only settles the net production or net injections that fall outside these contracts. However, membership of either system is compulsory for network users and the principles for pricing transactions in either market are the same. The difference between the two types of scheme can sometimes therefore be little more than cosmetic.

Settlement systems provide an ultimate source of electricity, available to all, whose centrally determined and regulated prices effectively put a cap on the value of electricity sold by other means. As we shall explain below, these prices often fail to provide the necessary incentives to build capacity.

The second problem is the targeting of outages. Customers who sign contracts with generators – directly or through retailers – cannot guarantee that spending more on capacity will necessarily improve their quality of supply. In the market paradigm, the market price rises to the Value of Lost Load when customers are being cut off, so that they are equally well off, whether they lose their supply or continue to take electricity. Both cases have the same, very high, cost as far as the customer is concerned. However, in practice, customers place different values on lost load, and pay a variety of contract and market prices, some of which may not reflect the true value of lost load during outages. As a result, customers are not in general indifferent between taking and losing supply of electricity.

In these conditions, investment in generation capacity, or security of supply, becomes a "public good" – i.e. a good or service that is available to everyone, regardless of the amount one pays for it.[24] Inevitably, everyone then tries to pay as little as possible for the good or service. To avoid inefficiently low levels of production, a public good or service must be provided centrally and funded by a compulsory charge or tax.

Box 2: The Public Good Characteristics of Security of Supply

Many commentators have simply assumed that security of supply is a public good, because it is impossible to prevent consumers taking power from capacity built and paid for by others. Such statements are true within a regulated tariff system, where the contribution that each customer makes to capacity has no effect on how often they are cut off. However, a market will only have public good characteristics, if the market price at times of shortage lies below VOLL. If the market price reaches VOLL during an outage, consumers who are cut off pay nothing towards the costs of capacity, whilst customers who receive a supply pay the full value of electricity, including a contribution towards the costs of capacity. Pricing at VOLL therefore removes the public good characteristics of security of supply - by charging the full cost of capacity to those who use it and preventing others from using capacity.[1] However, there are many reasons why electricity market prices might fail to reach VOLL during an outage, and which therefore reinstate the public good nature of security of supply.

3.2.2. Lack of demand-side signals

In most markets, only a small volume of demand enters the bidding process, so that it can help to set prices by the interaction of supply and demand. The bid price associated with some (or even much) demand has to be approximated by a rule, such as the imposition of prices equal to VOLL, when demand exceeds available capacity. However, these rules may rely on an estimate of VOLL which is too low (a matter of concern in Australia some years

[24] Examples of a public good are national defence or public parks. In both cases, one person's consumption of the service does not diminish the amount of the service available to others. Moreover, it is impossible to stop anyone from consuming (or benefiting from) these services if they don't pay for them. Together these two conditions define a public good.

ago) or the rules may systematically produce relatively low prices at all times (a matter of concern under NETA, the market arrangements in Britain). See Volume 2.

Furthermore, any market price may be subject to regulation that prevents it rising as high as VOLL, because price rises of this type are politically unacceptable or indistinguishable from abuse of market power. (Such price caps apply in many US markets, as well as in Spain.) However desirable such price caps may appear, they reduce the incentive to invest in capacity and require some offsetting mechanism.

3.2.3. Retail competition

Finally, it is becoming clear that retail competition has an effect on wholesale market operation and the level of investment. Many potential investors recognise that retailers facing competition cannot take on the risks associated with long-term fixed price contracts.

In SEE, the obligation to unbundle retail businesses and to allow retail competition will not affect countries outside the EU for some years. However, the same problem arises for integrated companies that own network assets and have a retail monopoly, if they are not operating on a sound financial basis. For instance, if regulated tariffs have been set below the level needed to cover costs, and if there is political resistance to increasing tariffs when costs rise, investors will not regard integrated utilities in SEE as creditworthy contract partners. This problem needs to be addressed before any incentives for investment can be effective.

Box 3: Consequences of Unbundling Retail Businesses

Integrated distribution companies possess network assets. They can therefore secure long-term contracts to buy electricity from generators against these assets. However, under the latest European directives, such companies are required to implement a “legal unbundling” of their retail and distribution network businesses. Such unbundling leaves the retail businesses unable to use network assets for security. If the wholesale price falls, these retailers would face competition from other potential suppliers able to buy low-cost supplies in the wholesale market. In such conditions, the existing retailers must lower their prices or lose market share – in either case risking bankruptcy. As a result, few generators will regard independent, competing retailers as creditworthy trading partners.

A recent paper analysed the effects of the transactions costs that prevent generators from signing long-term contracts with the thousands of final consumers of energy.[1] As might be expected, the effect of these costs is to increase the risks borne by investors in generation (or the costs of sharing the risk with consumers), and to reduce the level of investment that a market will induce.

Real markets incur further transactions costs when customers switch between retailers, and these costs discourage switching so that market shares are in fact more stable than might be expected. Retailers with a stable customer base may be prepared to take on some long-term contracts, thereby encouraging additional investment in generation. However, the lack of creditworthy counterparties remains a widespread problem for potential investors.

3.3. Price Spikes and Capacity Schemes

Energy-only markets face a problem if they provide price signals that result in extreme price spikes (i.e. short-lived – or even seasonal – jumps in electricity prices) and if these price spikes have unacceptable economic impacts when the market is short of capacity. Such conditions are most likely to prompt a regulatory intervention to cap prices, for the reasons given above, even though such price caps discourage the investment needed to alleviate the shortage. Capacity schemes are intended to avoid the problems that provoke regulatory interventions, or else represent a reaction to the problems caused by such interventions.

The purpose of such schemes is to make customers pay an insurance premium, for protection against outages and the price spikes that would otherwise occur when capacity becomes scarce. The insurance premium would pay for sufficient capacity to be on-hand, even if some of that capacity is very rarely called upon. Rather than recovering capital during periods of scarcity and very high prices, the capacity scheme seeks to avoid those periods and provide for a substantial portion of the capital recovery through the insurance premium.

3.4. Appraisal Criteria

If all the restructured electricity markets in the world had the same objectives, and placed the same relative importance on each objective, and if all the restructured electricity markets in the world had the menu of candidate capacity schemes available, then there would be a single "best" capacity scheme that all electricity markets would eventually adopt. However, in practice, a wide variety of capacity schemes exists and there is no single "best" solution. In practice, each market places a different relative weight on each objective, and each market has a unique set of options available. Accordingly, each market is led to a solution that is best for its own local conditions. In this section we discuss the objectives and the relative importance of each objective in the context of South East Europe.

The over-arching objective of any capacity scheme for SEE is to provide incentives, consistent with the energy markets in each member country, that encourage investors to provide an acceptable **security of supply**. In practice, this means investors must make enough capacity available to meet demand in all but extreme circumstances, and to do so efficiently.

What circumstances count as "extreme" depends on local conditions. The market paradigm tells us that an efficient plan of construction will allow more outages (i.e. more days when demand exceeds capacity) if capacity is more expensive relative to customers' valuation of lost load. The market paradism also tells us that efficiency requires an efficient mix of different capacity types, and an efficient pattern of use of existing capacity. Capacity can come from two sources: construction of new capacity at the right time and of the right type; and increases in the availability of existing capacity. Ideally, a capacity scheme would encourage the least-cost combination to these two sources. The aim of **efficiency** therefore has several dimensions.

In considering how to implement the scheme, it is important to maximise the scope for **competition**. In some cases, competition is a legal requirement (e.g. in retail supply), but it is also an important tool for achieving efficient (or least-cost) construction of new generation plant.

Further criteria help to elucidate the meaning of security of supply, efficiency or competition. Investors must have confidence that the promises contained in a particular revenue formula will be predictable and stable over the long-run, which requires that rules are as mechanistic as possible and not subject to change under political influences. All these features contribute to the **transparency** of long-term incentives.

To avoid political intervention in the markets, it is advisable to avoid price shocks to customers, either in the form of short-term tariff increases during shortages, or in the form of structural shifts in tariffs. Some of SEE's electricity industries charge tariffs below the level needed to cover present and future costs. Raising tariffs to improve cost recovery and to lower credit risk will be conducive to efficient investment. However, it will also be necessary to avoid consumer price shocks that undermine the public's faith in the system. **Consumer price stability** is therefore a desirable characteristic of any capacity scheme, even if it is defined in terms of a stable transition to tariffs that cover costs.

Finally, the detailed design rules must **rule out gaming** (manipulating the market) and meet general perceptions of **fairness**. Gaming is possible when the scheme rewards generators (or others) for providing false information (e.g. by saying that plant is unavailable, when it isn't, to drive up prices). *Within* any country, fairness can mean any combination of absolute prices levels, affordability, relative tariff levels (between different consumer groups) or non-discriminatory treatment of market participants. Whilst anticipating that such concerns will constrain the design of electricity markets, capacity schemes and final tariffs, we cannot anticipate exactly how such pressures will emerge in each country. However, we do anticipate that some concept of fairness will apply *between* SEE countries. Any capacity scheme must ensure that consumers in each country pay a fair share of the cost of investment needed to meet their own demand and a suitable reserve margin.

3.4.1. Summary

To summarise, capacity schemes are a substitute for high energy prices and are put in place to avoid or to overcome problems with an energy only market. Any capacity scheme must offer good overall performance by all the following criteria, with the emphasis placed in the first three:

1. Security of supply (adequate level of investment)
2. Cost efficiency:
   - appropriate risks to investors (credit risk, price risk, regulatory risk)
   - efficient long-term decisions on investment (how much capacity to build/ retire, what type of capacity and where and when to do it)
   - efficient short-term price signals (particularly, avoid distortions to dispatch)
3. Competition - ability of traders and consumers to choose who supplies them and on what terms
4. Transparency – not necessarily simplicity, but predictability and stability of rules
5. Price stability to customers

6. Non-susceptibility to gaming
7. Fairness

These criteria are not always straightforward to apply, but they do help to focus debate on the relative merits of different schemes.

4. Investment Incentives and Liberalisation

Given the difficulties of providing incentives for efficient investment, different electricity markets have adopted different devices for securing the investment they require. In this chapter, we review the main mechanisms designed to support investment in generation capacity.

4.1. Power Purchase Agreements

Power Purchase Agreements (PPAs) are a common investment mechanism in the South East Europe region and in many other parts of the world. The defining feature of the PPA model is usually that a single central authority (known under the old EU directive 96/92/EC as a "single buyer") buys power from selected Independent Power Producers (IPPs) using long-term contracts containing plant-specific terms and conditions. The single buyer may be responsible for buying energy for the entire wholesale market, or several single buyers may operate within one market, each one responsible for a single distribution network or some other part of the market. The PPA model is often used as a first step to liberalisation and as a way of attracting investment by private capital.

4.1.1. Prices in PPAs

The prices in PPAs are usually subject to some form of regulatory scrutiny, but are not always set equal to the generator's own cost of service. In large markets, the single buyer can hold some form of auction to determine the lowest cost offering. The buyer then signs long-term life-of-plant contracts with the winning IPPs after the regulator has approved the process of the auction and the selection of the winners. These contract prices are then passed on to the final customers as part of the bundled tariffs.

4.1.2. Efficiency and PPAs

The main advantage of the PPA model is that generators can be assured of long-term or even life-of-plant payments from a credit-worthy authority, and so generators can secure investment finance relatively easily and cheaply. This benefit should flow through to consumers via continued investment at lower cost, which should be reflected in a lower retail tariff (than would happen otherwise). This consideration is important in South East Europe where the investment risk in some countries is high by international standards.

It is possible to structure a PPA model to encourage efficient use of the power station, in a manner consistent with a short-term energy market, if the PPA adopts the common "two part" price structure:

- one "fixed payment" for fixed costs – i.e. for making capacity available; and
- another "variable payment" defined as a price per MWh of output and set equal to the plant's variable cost.

The efficiency of plant operation depends on the incentives created by the formula that defines the "fixed payment". This payment is only fixed in the sense that it does not vary with output, but it may vary in line with achieved availability, and incentives to make plant available may vary by season and by time-of-day, to encourage efficient decisions. There is considerable experience around the world with the design of such contractual incentives.

The separate definition of a “variable payment” allows the buyer to effectively simulate a market by calling on its PPAs in ascending order of the price per MWh of output. These prices can even form the basis for offering plant into a short-term market, which allows the market operator to identify a market price. This process works as long as the single buyer takes title to the plant’s output before it enters the market (i.e. at the station gate).

Another approach does not require such a change in title, but rather settles the contract in cash, using a Contract for Differences (CFD). Under a typical CFD, the IPP sells its output to the spot market, and the buyer buys all the electricity it needs from the same spot market. However, the IPP gives back to the buyer an amount equal to the difference between the electricity market price and the variable payment due to the IPP (based on its output or on a fixed volume of sales). The advantage of a CFD is that it does not require a complex process for notifying contract volumes to the system operator or the market operator, but can be settled bilaterally by the buyer and seller alone.

4.1.3. Competition and PPAs

This PPA model allows a limited form of competition. A market structure based on long-term contracts generally transfers market risk, technology risk, and most of the credit risk onto consumers, because the PPAs shelter the IPPs from market prices and from competition with improved technologies. The single buyer is generally a good credit risk, since it usually has a monopoly. (Hence, the PPA model can be an attractive model in countries that are perceived as having high investment risk.) These features mean that the IPP is deliberately shielded from competition once it has secured its contracts.

The competition in the PPA model occurs at the point where the single buyer selects the IPP that will be required to build and operate plants. The buyer selects the appropriate IPP by holding an auction and choosing the cheapest offer to supply. To make this choice possible, the buyer usually specifies a number of PPA terms, such as technology, fuel, and location, because of the difficulty of evaluating competing bids if the choices are wider. This predefinition of plant type limits the scope for competition, which often achieves efficiency by finding new technologies, fuels, and locations.

However, investors usually prefer or require the use of long-term (life-of-plant) contracts precisely because there are not enough buyers to permit full competition. In small markets, a generator without a PPA would be reluctant to “sink” large amounts of capital in a plant if he would then face only one potential buyer for his output (or even a small number of buyers). In such conditions, after the plant has been built, the IPP runs the risk of having to accept a price that just covers the plant’s running costs. To overcome this risk, new entrants into generation need to extract a commitment from the buyer to pay their capital costs, before they spend any money.

4.1.4. Disadvantages of the PPA model

While it is possible to establish a short-term energy market under a PPA model, in practice that short-term energy market is likely to contribute little to attracting private capital to the market. The single buyer has access to special arrangements for cost recovery, such as a (partial) monopoly over final consumers, or some additional source of revenue from networks. These special arrangements give the single buyer an advantage in trading, so that it will be prepared to invest in response to lower market prices than other investors. Indeed, in some systems (Ireland, Greece), it has been proposed that a central agency should sign contracts with new generators as soon as the market appears to be heading for a shortage. In such conditions, energy prices will be depressed by the new capacity contracted by the central authority and will not rise to the cost of entry by independents. As a result, the market becomes dependent on the actions of the single buyer to encourage investment.

In practice there is no major market using the PPA model in which the energy market prices are also signalling the need for investment. Hence, adoption of the PPA model could stifle the development of a meaningful power market.

Similarly, the PPA model could stifle development of the retail market. A retailer might have difficulty achieving the same creditworthiness as the central authority, and thus might have to pay a higher price for a PPA than the single buyer. This would make it difficult for the retailer to compete with the standard retail tariffs of the single buyer.

4.1.5. Credit risk and cost recovery

Retail tariffs in many South East Europe countries are currently below the level necessary to support new entry. The market paradigm assumes that wholesale prices rise for all generators when there is a shortage, and retail tariffs would have to rise accordingly to prevent bankruptcy of retailers. However, a PPA model does not require that the overall level of prices rise immediately to the cost of entry. Current tariffs may only cover low prices paid to existing generators, but need not rise by very much if the single buyer signs a higher price contract with a new entrant. The additional contract signals then need for new capacity, but does not raise the cost of supplies taken from existing generators.

Also, the PPA model is not prone to market power because generation is effectively still regulated by agreeing contract prices at the time the PPA is signed. This feature is important in some parts of South East Europe where the existing ownership structure gives some generators considerable market power. Under a PPA model, incumbent generators might be excluded from building new plants, so that ownership concentration would decrease over time, but in the mean time regulation would focus on giving the single buyer incentives for efficient operations and efficient conduct of its power procurement.

The PPA model provides a useful means of transition from central procurement to competition. A transition is attractive because it could allow time for retail tariffs to increase to the levels necessary to support new entry, and for market concentration to diminish.

4.1.6. Summary appraisal

As long as the contract is designed well, PPAs can provide a basis for encouraging efficient investment in and operation of power stations. The potential to use competition in selecting new generation provides a means for cost-minimising, although the desire for transparency usually requires the auction to take place within certain constraints on plant design. There is less potential for competition to serve wholesale and retail customers. It is possible to envisage a transition to competition, whereby electricity sold by PPA enters the spot market, but the effect of such reforms are limited unless the regulatory authorities are then prepared to let the market arrangements drive new investment rather than a system of centralised PPAs. Once a market has taken over, the value of each future PPA depends only on the value of electricity as defined by the market arrangements. PPAs then do not provide an independent mechanism for encouraging investment within a market, but only a means of sharing risk. The role of PPAs as an independent means of inducing investment is therefore a transition mechanism, available before any reliance is placed on the electricity market.

4.2. Energy-Only Markets

The preceding PPA discussion involved a model in which a central authority makes the decisions about building new generation. By way of contrast, under an energy-only market, all such decisions are left to decentralised market participants and they decide on the basis of energy price signals. The basis of an energy-only market is the signal provided by the charge for deficit imbalances when the market is short of capacity. In such conditions, the market must find a way to set the price of electricity equal to the value of lost load (VOLL) as defined by price-responsive consumers, or by a formula acting as a proxy for the value placed on electricity by consumers.

4.2.1. Problems with energy-only markets

The energy-only market model is problematic. The main problem is that in markets for electrical energy, customer demand tends to be unresponsive to short-term changes in wholesale price, and supply is also largely fixed in the short-term. This lack of flexibility on both the demand side and the supply side of the industry causes short-term prices to rise to very high levels at peak times. These price spikes are a problem for two reasons:

- First, as explained in section 3 and in Appendix A, in a functioning competitive market, generators cannot recovery their fixed costs if they do not receive the revenue from these price spikes. New generators will only enter a market if they expect to be able to recover all their costs. However, the revenue from these price spikes is highly uncertain there might be no spikes for a matter of years, or there might be a large number of spikes in a row when conditions of reserve scarcity arise (e.g. in a heat wave). There can be a great deal of uncertainty about how much generating capacity will be induced to enter the market given such a pattern of prices and the associated risk of financial problems if price spikes don’t occur. Accordingly, there can be a great deal of uncertainty regarding what level of security of supply will be achieved by an energy-only market.

- Second, the volatile prices are unattractive to consumers, regulators and governments because of the effect that volatile and unpredictable price spikes can have on individual consumers and on the economy of the whole. In an energy-only market, many regulators face pressure to impose price caps at levels that are “reasonable” by historical standards, but these price caps can have the effect of worsening the shortage that caused price spikes in the first place. If prices are capped at levels that are too low, it is even more likely that the market will not induce an acceptable level of security of supply. In Appendix A, we describe how, even putting uncertainty aside, prices need to spike to a level equal to the Value of Lost Load (VOLL). In Australia, VOLL has been set at AUS$10,000/MWh and similar or higher levels have been suggested in the United States. The experience in many countries has been that prices anywhere near these levels are unacceptable to consumers, regulators and governments – and particularly if prices could stay at these levels for more than an hour or two at a time.

The best way to make energy-only markets work is to make sure the market has sufficient demand-side price response, i.e. customers who can easily reduce their demand when prices rise. (There is not much that can be done about the supply side since it will always take at least two years or so to build a power plant). However, there is virtually no successful functioning electricity market in the world where demand-side response has been implemented in sufficient scale so as to provide the capacity solution by itself. Some markets have tried, and the notable example is California.

In California, the legislation and the initial design of the Californian market relied on demand response. However, the overall design of market and regulatory arrangements was flawed. The reforms included a freeze on retail tariffs for much of the state, so most retail consumers never felt the signals provided by wholesale prices. When wholesale prices rose above the level of the fixed retail tariff, distribution utilities made a loss and could do nothing to reduce it. As a result, the largest of these utilities were driven to bankruptcy and the California market failed. It is impossible to estimate what impact a capacity scheme might have had, but the lack of a coherent approach to rewarding capacity was a contributing factor nevertheless. California largely re-regulated its electricity market, and is now developing a form of capacity scheme that will be imposed on the utilities.

4.2.2. Energy-only markets in hydro-electric systems

Energy-only markets can operate more successfully in hydro-dominated systems, as opposed to largely thermal systems like California. Hydro systems that have storage (dams) tend to avoid many of the price spikes and associated investment problems because the supply side of the industry in a hydro system is more price-responsive in the short-term than in a thermal system.

In a hydro system there tends to be plenty of capacity available in any given hour; the constraining factor tends to be the amount of annual energy (water) available. If prices show signs of “spiking” in a given hour, hydro plants have the incentive and the ability to lift production in that hour above normal levels of production. This short-term flexibility in supply ensures prices in any one hour are similar to prices in neighbouring hours and avoids the short-lived price spikes that make investment so risky and provoke regulatory interventions.

Two of the best-known examples of energy-only markets, New Zealand and Scandinavia, operate in systems with a considerable amount of hydro storage. About 60-70% of New Zealand’s energy production is from hydro, whilst in Scandinavia hydro production accounts for 99% of production in Norway and about 50% of production in Sweden (with substantially less coming from hydro sources in Finland and Denmark). Nevertheless, both these markets have experience problems with prolonged periods of high prices due to a lack of hydro resources in "dry years".[25]

§ New Zealand has reacted to recent shortages by setting up a "dry-year reserve scheme" and by using government resources to promote investment in additional thermal generation. (See Volume 2.)

§ In Norway, hydro shortages have occurred intermittently during the last 10 years, and have brought about high prices for some winter seasons. These high price episodes have prompted much public discussion about the workings of the electricity market, but so far the government and the regulator have resisted pressure to intervene. Such decisions may be supported by the fact that hydro shortages in Norway, although they raise prices, still do not raise prices high enough or often enough to cover the long-term cost of building new generation capacity. (See Volume 2.)

In South East Europe hydro provides a substantial proportion of total energy production in some countries. However, much of this hydro capacity is run-of-river, with limited storage capacity and, hence, limited ability to smooth out prices and to avoid price spikes. As a result, it is likely energy-only markets in South East Europe would be prone to (and have to rely on) price spikes to remunerate investment in capacity. In any case, it will be some time before regulatory and political conditions in the region can tolerate a large rise in wholesale market prices and its effect on consumer tariffs.

4.2.3. Summary appraisal

Although the market paradigm implies than an energy-only market can provide efficient investment incentives, it is widely acknowledged that the incentives may not produce either a level of security or a level of prices that is consistent with political and regulatory constraints. It may be considered incompatible with fairness that generators should earn their highest profits at times when consumers are being forcibly cut off (even though such profits are necessary to guarantee the achieved level of security of supply).

In any case, it is almost impossible to distinguish between genuine shortages and a shortage caused by generators withdrawing plant during periods of high demand. Such suspicions of market manipulation are enough to provoke many regulatory investigations, and even the threat of such investigations can undermine investors' confidence that they will be allowed to earn profits from high prices.

The adoption of an energy-only market is therefore a relatively simple option in conditions of excess capacity, when prices are low. Many of these markets have been created precisely in these conditions (Norway – 1991, New Zealand – 1996, California – 1996, NETA – 2001), but now have (or will soon have) to face up to the implications of energy-only pricing in conditions of shortage. The lessons from these countries suggest that electricity markets in SEE are vulnerable to many of the problems that have prevented energy-only markets working smoothly in other countries. Thus, although the energy-only market has many

[25] In New Zealand, a dry year follows a lack of snow- or rainfall over the previous winter. In Norway, a hydro shortage may occur during a cold winter, when a lack of snowfall (and an anticipated future shortage of production) coincides with a cold or extended winter (which raises the current level of demand).

potential advantages, it has known short-comings, and seems inappropriate, in the conditions found in SEE.

4.3. Capacity Price Schemes

The main alternative to a PPA program or an energy-only market is some form of capacity scheme operating in conjunction with an energy market. There are two major forms of capacity scheme:

1. Schemes that fix a capacity *price* (or price formula) and make payments at this price to all generators; and
2. Schemes that set a *quantity* of capacity required to ensure adequacy and impose an obligation on either a single buyer or on all retailers to provide (build or buy) such a quantity.

The two schemes can share many common elements, but are fundamentally different. This section discusses capacity *price* schemes, and the section that follows discusses capacity quantity schemes.

4.3.1. Outline of scheme

The key elements of capacity *price* schemes are:

§ A price is paid to available capacity, on the basis of some formula;

§ A method is established for collecting the funds required to make these payments from distributors/ retailers or end-use customers;

§ Investors allow for the revenues expected to derive from the capacity price when deciding whether, or when, to invest in new capacity (or to shut down existing capacity); and

§ The market determines what quantity of capacity will be supplied, given the established price.

The transparency and predictability of capacity payments under this scheme induces capacity to be constructed or maintained that would not have been constructed or maintained under an energy-only market, because it reduces the risk associated with maintaining or constructing new capacity. On one hand, capacity remuneration no longer depends on the level and frequency of occasional price spikes. On the other hand, the avoidance of price spikes makes it less likely that regulators will face pressure to cap prices below the level needed to remunerate capacity.

4.3.2. Eligibility for the capacity price

A variety of mechanisms can be and have been used to fix the capacity price. The key attribute of a fixed capacity price is that either a price or a formula is specified by a central entity with authority over the market. The payment is then made to generators who provide the capacity product. The payment is intended to *supplement* energy payments. As explained earlier, the payment is justified on the grounds that the energy market might not provide the level or stability of prices needed to encourage the development of sufficient capacity in a timely enough manner to ensure supply adequacy. This may particularly be the case as a result of limits on prices due to practical (IT) reasons, price caps, or energy market mitigation schemes that keep energy prices near the marginal costs of production even during periods of capacity scarcity.

As described in Volume 2, the capacity scheme originally used in England & Wales set a capacity payment for all generation that was calculated half-hourly, and was based on day ahead projections of half-hourly loss of load probability (LOLP) and a value of lost load (VOLL) established by the regulator. This calculation had two components, which complemented one another. First, the energy price for day-ahead sales to the Pool includes a "*capacity adder*". Second, a generator that offered but had not sold energy day-ahead to the Pool was granted a separate (and broadly equivalent) *"availability payment"* to induce it to declare plant available at the day-ahead stage, and to keep it available in real time.

The payment to generation in the England & Wales pool took two different forms:

Revenue of capacity required for day-ahead sales

= SMP (for day-ahead sales) + capacity adder (for day-ahead sales)

= SMP + LOLP*(VOLL-SMP)

Revenue of capacity *not* required for day-ahead sales

= offer price (for real-time output) + availability payment (for capacity)

= Offer price + LOLP*(VOLL-offer price)

In the second formula, the offer price for the capacity *not* required at the day-ahead sales is only paid if the capacity actually generates the output. If the output is zero then the capacity gets only the availability payment based on the offered level of output.

Note that the scheme required a definition of VOLL, just as in some energy-only markets, and that the payment effectively included a deduction when the plant receive a high energy price (SMP) or offer price.[26]

If the Pool had not paid out the "availability payment", generators would have had a strong incentive to offer their plant more cheaply, to make day-ahead energy sales and to earn the "capacity adder". The result would have been a reduction in day-ahead energy prices that offset the beneficial effects of the "capacity adder" and distorted the pattern of plant dispatch. Thus, any scheme of "capacity adders" needs to be supported by payments to other plant.

Ideally, capacity payments should accrue to generation capacity whether it is running or not, in order to avoid this kind of distortion. The definition of capacity that is eligible to receive the payment might take into account differences in reliability (especially if it is based on day-ahead or even year-ahead calculations). But it would definitely be inefficient to tie capacity payments closely to actual levels of output. Most capacity payment schemes require

[26] See Volume 2 Box 1 page 14, for an explanation of the rationale for these formulae.

occasional or continuous tests of availability, to ensure that only functioning capacity receives the payment.

4.3.3. Capacity formula (1): LOLP

Ideally, the price of capacity should rise when it is in short supply and should fall when it is in excess. In the old Electricity Pool of England and Wales, the capacity adder was responsive to the quantity of capacity available, because LOLP was a function of available capacity. Indeed, many capacity price schemes do have some sort of link to the quantity of capacity that exists in the market. The key point that distinguishes a capacity price scheme from a capacity quantity scheme is that, while quantity might enter the formula in setting the price, ultimately it is left for the market to decide on the quantity of capacity to enter or leave the market, given that price.

In the price formulae used in the England and Wales Pool, the LOLP element produced additional revenue that depended on the state of the market, in particular on the capacity margin: the lower the level of capacity, the higher the value of LOLP, and the greater the reward for capacity (whether by keeping existing plant available, or by building new plant).

LOLP, and the associated revenue, proved volatile and difficult to predict. Investors usually sought out other market participants to sign contracts with which they might stabilise their revenues. Furthermore, the payment was subject to manipulation by generators withholding small amounts of capacity. As a rule of thumb, a reduction in capacity of 500 MW (roughly 1% of peak demand) caused LOLP (and the value of capacity) to double. In practice, the large generators probably used this market power to stabilise total prices over the year as a whole, but the perception that the market was open to manipulation did not enhance incentives to invest.

Absent gaming concerns, however, such a formula adds to the energy price an amount equal to the expected value of load that may not be served in each half hour period, and is considered to provide an efficient short term price signal. The rules adopted for England & Wales applied a degree of smoothing, to limit half-hourly volatility but, taking a year at a time, revenue from this element of the price varied considerably and such a formula might not be consistent with the objective in South-Eastern Europe of limiting exposure to spot prices that are volatile and may be unacceptably high. Generators would need to forecast the highly uncertain level of revenue from such a scheme for use in investment decision making.

4.3.4. Capacity formula (2): Inverse relation to available capacity

A capacity price scheme can also be designed so that the capacity payment to each unit of generation capacity is set for a period of time and announced in advance. This variant of a capacity price scheme is referred to as a ***price signal*** method in this paper. Eligible capacity receives the established capacity price. A methodology is needed to determine the price, and the timing and method for resetting the price also need to be specified.

An existing example of such a scheme is found in the Spanish market. In Spain, the overall level of capacity revenue is fixed in advance, as a charge on all consumption defined in €/MWh. This revenue is allocated among generators on the basis of their individual capacities at peak times. As a result, when capacity is short, generators receive a higher capacity payment per unit of capacity. When capacity is in excess supply, more generators must share a fixed revenue and the capacity payment per unit of capacity falls. This kind of formula provides a transparent means of adjusting the payment for changes in supply conditions. In practice, however, it has proven difficult to link the charge on consumers to any objective factors, so capacity payments have proven somewhat unpredictable. Ideally, the scheme would estimate the required charge on the basis of forecast demand, peak capacity requirements and the cost of peaking capacity.

4.3.5. Summary appraisal

Capacity payment schemes can provide relatively efficient incentives for plant construction and operation, but their effects depend crucially on how the level is defined and who is eligible to receive them. Some of the schemes have proven unpredictable, either because the rules for setting the payment were not clear, or because the scheme was open to manipulation (real or threatened) by large generators. Such problems undermined both the incentive properties of the scheme and its perceived fairness as a way of charging consumers for investments.

One problem that arises with capacity payments is the difficulty of using a price to achieve a particular volume of construction. If the price is set too low, there will be too little capacity, and vice versa. Both the England and Wales scheme and the Spanish scheme adopted an automatic adjustment formula, so that the price of capacity rose and fell in response to the market situation. Such adjustments reduce the impact of errors in estimating the required level of capacity payments.

4.4. Capacity Quantity Schemes

The second broad alternative type of capacity scheme is that a central authority establishes a *quantity* of capacity that is needed, and the market determines the price. Such schemes are also known as “capacity obligations”, because the requirement is imposed on one or more market participants as a regulatory obligation.

4.4.1. Basic design of the scheme

The required quantity of capacity is most often established by analysing Loss of Load Expectation (LOLE). This analysis looks at the volatility of demand and the reliability of generator capacity, to determine how many hours a year on average demand is likely to exceed capacity. The regulator must set (or approve) a desired standard, such that load should be lost less than X hours per year on average. The required amount of capacity would be the level that results in an expected LOLE of X hours per year.

Having identified the required “installed capacity” (ICAP), electricity markets allocate this obligation to “load serving entities” (i.e. retailers). The markets can then adopt either of two general methods to achieve the required level of capacity:

§ **Type 1:** Retailers can be obliged to build or buy sufficient capacity to fulfil their share of the overall obligation and will incur a penalty if they fail to procure enough capacity.

§ **Type 2:** A central authority can procure such capacity (but not energy) through a competitive process such as an auction or an “Invitation to Tender” (ITT). The central authority then allocates the costs of these purchases to retailers in proportion to their obligations. In most schemes, distributors or retailers can avoid this cost by building or buying capacity independently, but if they don't they are allocated a share of the cost resulting from this competitive process, to the extent of the shortfall between their own capacity and their assigned obligation.

In practice, the only substantive difference between these schemes is whether a retailer who is short of capacity pays (Type 1) a financial penalty (which encourages the retailer to seek out capacity) or (Type 2) the cost of actually procuring capacity through a central auction. Under either method, the capacity obligation might be updated as frequently as daily, although in practice it might be updated as often as consumers can change retailer.

Since generators sell capacity either to distributors/ retailers or to the central authority, capacity quantity schemes provide generators with a source of revenue that helps induce capacity investment and ensures that sufficient capacity is available.

4.4.2. Detailed design questions

Volume 2 contains detailed descriptions of capacity obligation schemes that are operational or proposed in the US and in Greece (a member of the SEE regional electricity market). The basic elements of each scheme are:

1. The definition of each retailer's obligation as the sum of its forecast peak demand and the required reserve margin;
2. The definition of capacity eligible to count towards the obligation, including arrangements for trading capacity between different markets or (in some cases) between different zones within one market;
3. the process for coordinating maintenance outages;
4. The process of monitoring compliance on a daily or monthly basis, including arrangements for recalculating obligations when customers switch between retailers; and
5. The penalty charged on retailers whose holding of eligible capacity falls short of their obligation.

The most serious design problems arise out of the need to provide a value of capacity that reflects the high implicit value of energy at peak times. Some schemes have had to operate under capacity price caps, or with a relatively low penalty for a capacity shortfall. The consequence may not be a lack of investment, but rather a distortion in the pattern of trade. Generators try to sell their energy to the market with the highest penalty for a capacity shortfall, rather than the highest true value of capacity. This problem indicates a need to coordinate market design over a wide area.

The arrangements for trading capacity between zones or markets provide a means for sharing capacity efficiently, when different markets experience peak demands at different times. Trading is therefore to be encouraged, when it follows a rational path, but requires the rules on eligible capacity to define what transmission access a generator in market A must have before he can sell capacity in market B.

There is widespread agreement that the market needs to monitor capacity obligations in advance (i.e. on the basis of forecast demand, not current demand), to avoid price volatility in the market for capacity. Given that capacity schemes are intended to avoid reliance on volatile energy prices, the goals can be undermined if the market price of capacity is also extremely volatile. In real time, the value of capacity is either derived from VOLL (when it is in short supply) or else zero (when there is sufficient capacity).[27] However, the ***expected*** value of capacity at some future date varies less. As a result, checks based on forecast demand for the coming year produce a less volatile market and more sustainable prices.

Finally, an important question concerns the obligations placed on generators who sell their capacity to retailers. Such generators may face a penalty if they are not able to provide enough capacity to back these sales. However, in addition, many capacity obligations require generators to offer their capacity to the system operator, so that its output is available in the real-time spot market or balancing mechanism. This obligation on generators gives greater confidence that capacity will be available and used in an emergency.

4.4.3. Perceived problems

Recent innovations in capacity market design address two perceived problems. First, when generators received high prices in spot markets during capacity shortages, customers and regulators complained that the generators were earning too much and that customers were getting nothing in return when they paid for capacity. In fact, the capacity obligation may have encouraged a higher level of investment and reduced the number of episodes when generators earn high prices. However, rather than impose a price cap on energy prices, some regulators have decided to adopt a remuneration scheme similar to a Contract for Differences (CFD).

To fulfil their capacity obligation, retailers must build capacity or buy capacity through a standard contract. Associated with this standard capacity contract is an implicit “strike price”. When the energy market price rises above this strike price, generators must rebate the difference to the holders of their capacity contracts. In this way, the capacity contract obliges the retailer to pay for capacity, but relieves the retailer of the burden of high energy prices.

The second problem arises when capacity prices are volatile, as unpredictable revenues do not provide efficient signals for investment. As a result, just as the designers of capacity payments decided to make them responsive to the market situation, so the designers of capacity obligations have recently decided that the penalty for a capacity shortfall should adapt to system conditions. As a result, the design of capacity obligations is moving closer to the design of capacity payments; we describe these innovations below under the heading of “hybrid systems”.

4.4.4. Summary appraisal

Capacity obligations defined in quantity terms have been a feature of centrally planned or coordinated electricity systems for many years. The introduction of wholesale and retail

[27] See Volume 2, for a discussion of experience with a real-time capacity monitoring system in New England and PJM (Pennsylvania-New Jersey-Maryland) .

competition required the adoption of systems to transfer obligations as customers switch between retailers, but did not undermine the operation of the basic scheme. However, retailers whose prices are set by competition (or by regulations that pass through the cost of buying energy and capacity in wholesale markets) are exposed to the risk that capacity prices fluctuate unpredictably, leading to high price episodes and a need to raise consumer prices. These episodes create the same problems as afflict energy-only markets during a shortage. To solve the problem, designers have adopted a number of schemes intended to smooth out such fluctuations, including the imposition of obligations in advance, based on forecast demand.

In principle, capacity obligations have the power to provide efficient signals for producers to construct an efficient quantity of capacity, but the efficiency of the scheme depends upon the accuracy of methods used to estimate capacity requirements, on the level of penalties for non-compliance, and on the potential for distortion of trade between different markets for capacity.

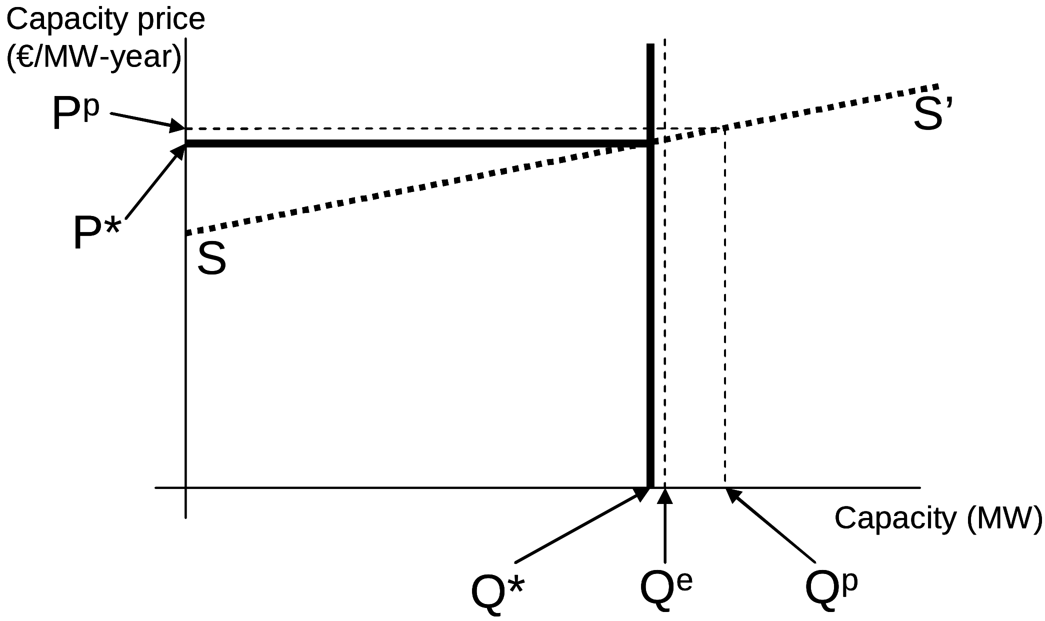
The relative merits of capacity obligations over capacity payments therefore depend on local conditions and the details of the design.

4.5. Hybrid Schemes – Demand Curve for Capacity

Capacity price schemes set a price and let the market determine a quantity – the quantity supplied by the market is an uncertain variable. By way of contrast, capacity quantity schemes set a quantity and let the market determine the price – and the price determined by the market is an uncertain variable. Each scheme has the disadvantage that while one factor is known with certainty, the other is highly uncertain.

Figure 4.1 shows the potential errors that arise under each scheme, assuming (1) that the supply of capacity (the line marked SS') is relatively flat, meaning that the cost per unit of new capacity does not rise much if the volume of investment rises and (2) that the demand for capacity is fixed (the vertical lines). In this figure, the efficient level of capacity is Q* and the equivalent market price for capacity is P*.

Figure 4.1
Quantity and Price of Capacity



One must anticipate some uncertainty over the estimate of capacity prices or requirements, and Figure 4.1 shows the consequences of different types of error. For example, if the central authority overestimates capacity requirements by a small percentage and sets it at Q^e instead of Q*, customers must pay for more capacity, but the price of capacity does not change much. (Indeed, the increase is too small to show in the figure.) However, if the central authorities tries to set a price for capacity and overestimates it by a similar small percentage, the consequence is a large rise in the volume of capacity that investors are prepared to build and hence a large rise in costs to consumers.

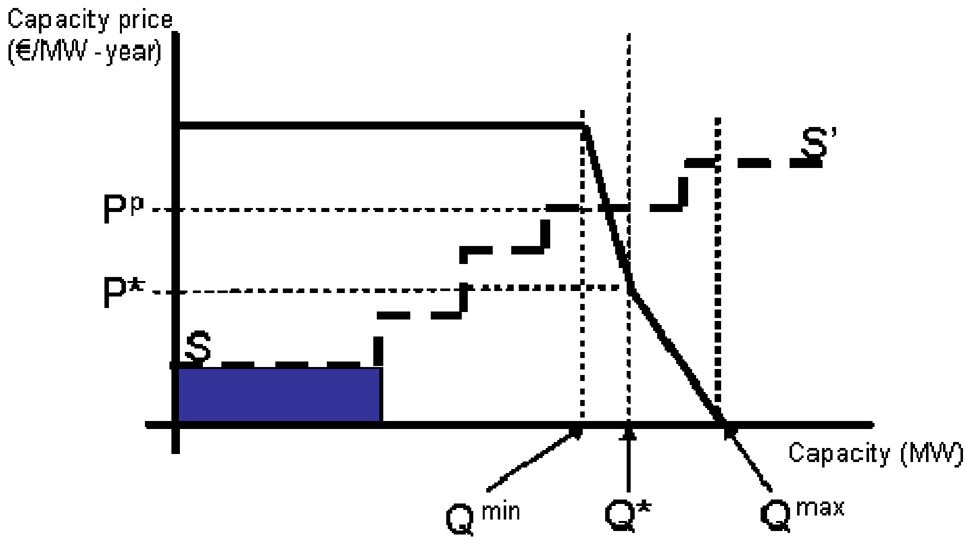
This analysis indicates that, in conditions of uncertainty, a capacity obligation is likely to produce a more efficient outcome than a capacity payment. This result derives from the flatness of the supply curve (SS') and the steepness of the demand curve (i.e. the obligation). These conditions are near universal.

Hence, if faced with choice between these two capacity schemes, countries in the SEE region might decide that the highest priority objective is to adopt a capacity scheme whose outcome is a predictable level of capacity sufficient to provide an acceptable level of security of supply. This appraisal criterion appears to point towards a quantity scheme being the better choice. However, in SEE conditions, it is not at all certain what level of security customers are willing to pay for, particularly since the costs of building capacity in each country may differ considerably, due to differences in risk, fuel availability, local costs, etc. Setting the same reserve margin obligation in each country might have unfortunate consequences:

- In one country, meeting this reserve margin obligation might require construction of extremely expensive capacity, which no consumer would be willing to pay for;
- in another country, the same reserve margin might achieve a lower level of security than customers want.

There is a way to dampen the effects of such errors, by allowing the required level of security to adapt to information about the actual costs of capacity. New England operates a "type 2" capacity obligation, in which the Independent System Operator (ISO) runs an auction to buy the capacity that retailers need to meet their obligation. The ISO is now proposing to adopt a "downward-sloping demand curve" for capacity in this auction, rather than buying enough to meet a fixed obligation. Figure 4.2 shows how the scheme is intended to work.[28]

Figure 4.2: Hybrid Scheme with Downward-Sloping Demand Curve



In this diagram, Q* represents the desired level of capacity. This figure can be taken from standard industry practice, but the system designers claim to have derived the appropriate level by modelling a choice between building capacity (at the anticipated cost of a peaking generator, P*) and losing load (which costs the Value of Lost Load). Q^{max} is then the level of capacity at which the price of capacity falls to zero, beyond which further investment has no value, whilst Q^{min} is a minimum tolerable level of capacity, at which the value of capacity is set equal to 2 times P*. (For instance, Q* might represent a reserve margin of 18%, whilst Q^{max} is equivalent to a reserve margin of 30% and Q^{min} to a reserve margin of 12%.)

The shaded area represents the capacity that generators have already sold to retailers and the curve SS' shows additional offers that generators submit to the ISO, to sell particular volumes of capacity (in MW) at different prices (in $/MW-year). The ISO then buys capacity up to the point where the supply line SS' crosses the demand curve, which in this case is at a level of capacity between Q^{min} and Q*, at a price of P^p. The ISO then works out which retailers are short of capacity and allocates the cost of its purchases to them. The price resulting from this auction therefore represents both the market price of capacity and the penalty that retailers pay for being short of capacity

This process means that, if the price of capacity rises, the ISO and retailers are obliged to procure a little less capacity. The parameters of the demand curve have proven difficult to

---

[28] The following explanation comes from P. Cramton and S. Stoft (2005), A Capacity Scheme That Makes Sense, 25 March 2005, available at http://www.cramton.umd.edu.

define objectively (although an explanation is possible), but the scheme results in the security of supply delivered to customers in New England area adjusting for variations in the cost of buying capacity.[29]

4.6. Conclusion

This discussion of possible capacity schemes has highlighted both the need for some method of rewarding investment in generation capacity and the different characteristics of each scheme. It must be noted that there is no universal agreement on "the" best capacity scheme for electricity markets and that different schemes work better in different conditions. The following section applies our evaluation criteria to each of the schemes, for the conditions applying in the SEE region.

[29] The calculation of the downward sloping demand curve is apparently complex, and difficult to justify objectively, but in fact it approximates to the LOLP function. The ISO estimates a capacity requirement (peak demand plus reserve margin of 18%) and sets the capacity price at that level equal to the annual cost of peaking capacity. Then, the price of capacity falls linearly to zero if capacity rises above the required level by 15% and doubles if capacity falls more than 6% below the required level. These parameters have been set so that, assuming some random variation in the level of capacity, average long-run revenues match the annual cost of peaking capacity. See Cramton and Stoft (2005).

5. Formal Appraisal of Capacity Schemes

As each of the schemes must be appraised against the alternatives, we consider all the schemes together, by the appraisal criteria listed in section 3.4, as applicable in the conditions of SEE. The possible grades are 1 (= poor), 2 (= unsatisfactory), 3 (= neutral), 4 (= good) or 5 (= excellent).

5.1. Security of Supply

We appraise the various schemes as follows, with respect to their ability to encourage enough investment for security of supply:

§ PPAs (Single Buyer) – good prospects, provided that the single buyer is creditworthy now and in the future, taking into account the regulation of its revenues and the prospects for competition; score = 4.

§ Energy-only markets – a few successes in developed markets, but particularly dependent upon the prices defined in the market during periods of shortage; investors tend to be wary of relying on infrequent and short-lived price spikes, because of the financial problems any project faces during the periods in between; and because of the possibility of intervention to cap prices; works best with demand-side response, which tends to be underdeveloped in SEE; score = 2.5.

§ Capacity price schemes – again, some apparently successful cases, but the degree of success depends on the accuracy of the estimated cost of capacity, which is not easy to estimate in SEE conditions and which could therefore lead to sustained over- or under-investment; score = 3.

§ Capacity quantity schemes – reduce the problem with misestimating the correct value, since there is relatively little variation in desired reserve margins, even when the cost of capacity is not know with certainty; the success of the scheme depends exclusively on the level of the penalty imposed on companies that do not contract for enough capacity; score = 3.

§ Hybrid Schemes – largely untried system, but combines the advantages of price and quantity schemes, and builds in an error correction mechanism to allow for errors in estimating the cost of capacity; score = 4.

As discussed above, each of the schemes described above has the potential to encourage an adequate level of investment, but there are major concerns about the success of each scheme in practice. Hence, the scores are very close to one another and selection of a scheme will depend upon other factors.

5.2. Cost Efficiency

The different schemes have very different implications for the efficiency with which investors handle (1) risk, (2) long-term decisions and (3) short-term decisions.

§ PPAs – The purpose of a PPA is to pass price risk to the buyer, which greatly reduces the price paid for the power (but exposes customers to long-term risks). The efficiency of long-term decisions about capacity type depends on the design of the capacity auction or selection process, which limits innovation and may be amenable to influence by interest groups. Fixing prices for long-periods will encourage investors to cut costs, but buyers only benefit at the selection stage if cheaper suppliers bid lower prices. The short-term efficiency of operations depends on the incentives in the contract. Score = 2.5.

- Energy-only markets – If well designed, a market provides the most efficient signals for risk management, for long-term efficiency and (particularly) for short-term efficiency, but only if supported by long-term contracts. Abstracting from the problems with energy-only markets in SEE conditions (which are discussed elsewhere), this approach scores highly, as its sets the efficient standard; score = 5.
- Capacity price schemes – As long as the formula reacts to market conditions, it provides better signals than a PPA, but worse than an energy-only market; score = 3
- Capacity quantity schemes – similar to capacity payments, but with more scope for markets to determine outcomes, score = 3.5.
- Hybrid Schemes – dominates other capacity schemes, because of the price-responsiveness of demand for capacity; score = 4.

5.3. Competition

The different schemes allow different amounts of competition, defined as the ability to choose.

- PPAs – This scheme only allows competition at the stage of selecting a power source. Competition at that stage may capture most of the benefits in terms of cost minimisation, but does nothing to foster innovation in wholesale and retail markets (or even in production, if the contract form limits the choice of technology). Creating competitive conditions later can be difficult, as it requires divestment or contract sales and some treatment of stranded costs. Score = 3.
- Energy-only markets – Once again, this is the definitive system for competition and would normally score 5. In principle, energy-only markets support competition among generators and among traders and retailers. Such markets have many supporters on this basis alone. However, in SEE conditions, competition is likely to be limited by a variety of interventions and government support mechanisms. In some member countries, there will be limited competition at the retail end. Score = 4.
- Capacity price schemes – This scheme allows buyers to choose how they buy electricity and so permits decentralised competition to buy and produce electricity. However, the payment for capacity is determined centrally and may be over- or under-estimated. Score = 3.
- Capacity quantity schemes – These schemes require central determination of the appropriate quantity of capacity, but give freedom to buyers to choose when and how they buy it and energy. The price of capacity is determined by competition in the market for capacity, although it may be volatile. Score = 3.5.
- Hybrid Schemes – If demand for (i.e. the obligation to buy) capacity responds to price, the system operates even more like a market and allows further competitive choices over price or security. Score = 4.

5.4. Transparency

Transparency is defined by the predictability and stability of the market rules, which is a tough standard to meet in the context of new electricity markets and underdeveloped legal frameworks.

- PPAs – Investors like PPAs because they provide a chance to define clearly the obligations and rights of both sides. Lack of clarity in drafting, or in the legal framework for dispute resolution, can undermine this predictability. Score = 4.
- Energy-only markets – Investors have shown interest in energy markets when they have operated on a stable basis for many years, without intervention by government. However, fears over intervention to stop prices rising (including not just price caps but also government-sponsored investment that keeps prices low) undermines transparency and investor confidence. New markets, such as those in SEE, start with this disadvantage. Score = 3.
- Capacity price schemes – These schemes provide a more stable revenue for capacity than energy-only markets, but rely heavily on administrative decisions about the price itself. Score = 3.
- Capacity quantity schemes – Definition of the required quantity is usually a technical process that produces similar answers in all cases. However, lack of clarity over the basis for penalising market participants who are short of capacity undermines the effectiveness of the schemes. Score = 4
- Hybrid Schemes – Whilst these schemes have many theoretical benefits, there is not enough experience with them to determine how to set the demand curves for capacity, Hence these schemes suffer from a lack of transparency over these parameters. Score = 2.

5.5. Price Stability to Customers

This criterion is important as it affects the long-term sustainability of the scheme and hence the transparency of incentives and the potential to foster investment in security of supply.

- PPAs – This method provides long-term stability, both because costs to consumers are fixed in the PPAs and because *average* prices only rise by a small amount when the single buyer has to start paying for new capacity. Prices are therefore very stable (although that also means that retail prices don't signal the true marginal cost of consumption). Score = 4.
- Energy-only markets – Such markets have proven to be very unstable. With retail competition, long-term variation in wholesale prices feeds through into retail prices. With regulation of retail prices, even short-term variation affects retailer customers unless the retailer is obliged to take on contracts to limit such risk. Score = 1.
- Capacity price schemes – Such schemes provide a moderate degree of stability, because the price for capacity is fixed. Score = 3.

- Capacity quantity schemes – Such schemes allow the price of capacity to vary and are therefore subject to some instability, unless the obligation requires the retailer to take on fairly long-term contracts (e.g. a year or more). Score = 2.
- Hybrid Schemes – The purpose of introducing price response is to reduce the effect of errors in forecasting the required payment to capacity, and thereby to reduce variations in the total cost to consumers, compared with a fixed capacity payment. Score = 3.5.

5.6. Non-Susceptibility to Gaming

This is a hard criterion to appraise, because the potential for gaming depends on the detailed rules of any scheme, which may vary from place to place and time to time. As a result, the scores do not vary much between schemes.

- PPAs – Long-term contracts prevent generators from exercising market power and are therefore useful in concentrated electricity markets, such as those found in some parts of the SEE. Holding tenders for such contracts prevents incumbents from using market power. However, using auctions to buy short-term power only from incumbents will reinstate generators' ability to use market power. Negotiation processes create opportunities for gaming (either by the buyer or by the seller). Score = 4.
- Energy-only markets – All markets are vulnerable to market power and energy markets are no exception. Problems arise at times of peak demand, when it is difficult to distinguish between genuine capacity shortages, and market manipulation. Score = 3.
- Capacity price schemes – Assuming that the capacity price is fixed and the energy price is capped, generators lose money by withdrawing capacity at peak times. Hence, there is less scope for manipulating market prices at peak times (though it remains at other times). Score = 4.
- Capacity quantity schemes – Some schemes have faced accusations that generators are manipulating capacity markets by withdrawing capacity to create a shortage. In these schemes, such actions would drive up the price of capacity. Score = 3.
- Hybrid Schemes – The introduction of price response in the demand curve for capacity puts this scheme between capacity price and quantity schemes. Score 3.5.

5.7. Fairness

As discussed above, it is difficult to appraise the fairness of a scheme by objective means, so we award all schemes a score of 3 (= neutral). In passing, we note the following:

- PPAs – have been subject to ex post criticism about the terms on which they were offered or the prices received by investors, and so are sometimes criticised as unfair on consumers;
- Energy-only markets – subject to many criticisms, especially if prices reach their highest levels at times when some customers are being cut off;
- Capacity price schemes – consumers may criticise the payment of a fixed price, on the grounds that it "exempts generators from true market pressures" (as, indeed, it is intended to do, so that consumers don't suffer wide variations in price);

- Capacity quantity schemes – the idea that retailers should contract *in advance* for capacity is not subject to the same criticism that generators don't face market pressures, but if the obligation appears to be overstated or expensive, some consumers will complain that they are paying too much for security that they don't need or want.
- Hybrid Schemes – try to overcome the criticism that security is too high or too expensive by linking price and quantity; derivation of the demand curve is still an area of concern.

5.8. Summary

Table 5.1 shows the scores awarded above, the unweighted total, a weighted total (and average) and the ranking of each scheme's score in both cases. On an unweighted basis, PPAs and the hybrid capacity scheme (with price response) both come out on top, with the others lagging behind. On a weighted basis, the hybrid scheme pulls ahead of PPAs, but the others still lag behind.

Table 5.1
Summary of Formal Appraisal

Criterion		Weight	PPAs	Energy only	Cap Price	Cap Quantity	Cap Hybrid
1	Security of supply	5	4	2.5	3	3	4
2	Cost efficiency	4	3	5	3	3.5	4
3	Competition	4	3	4	3	4	4
4	Transparency	3	3	3	3	4	2
5	Price stability to consumers	2	4	1	3	2	3.5
6	Non-susceptability to gaming	2	4	3	4	3	3.5
7	Fairness	1	3	3	3	3	3
	Unweighted Total	21	24	21.5	22	22.5	24
	Ranking		**1**	**5**	**4**	**3**	**1**
	Weighted Total		72	68.5	65	70	75
	Weighted Average		3.4	3.3	3.1	3.3	3.6
	Ranking		**2**	**4**	**5**	**3**	**1**

The difference between the weighted and unweighted scores lies in the extra weight put on cost efficiency and competition. With respect to these criteria, PPAs perform less well than a capacity scheme that decentralises both capacity obligations and trade in energy. This implies that PPAs may be preferred in systems that are not required to open up their markets to competition, but become less preferable once the creation of a competitive market becomes a priority.

The discussion above leads to some clear conclusions about the potential for each scheme in SEE conditions:

- Power Purchase Agreements: as a central means of procuring all new capacity, the use of PPAs has had some success in SEE and elsewhere. However, the success is conditional on investors believing that the buyer has the means to recover its costs and hence to meet its obligations to the generator. Moreover, a generalised system of centrally procured

PPAs is no longer consistent with EU directives on the internal electricity market, since its effect is to hinder competition in wholesale and retail markets. PPAs can provide a transitional system but it is necessary to anticipate the introduction of competition in wholesale and, eventually, retail markets. Some SEE countries will face this challenge sooner than others (with Greece already facing it).

§ Energy-Only Markets: although the market paradigm suggests that energy-only markets will provide good incentives for investment, in practice they suffer from many constraints. The paradigm shows that prices would have to rise to very high levels during hours of lost load – or else there would need to be a very high volume of price responsive demand. Neither condition is likely to hold in SEE, given the observed problems in other markets. If the market rules cause prices to rise sufficiently during shortages (a matter of concern in Britain), regulators may nonetheless feel pressure to cap prices to prevent market manipulation (a common feature of US markets and an underlying threat in New Zealand). The theoretically superior efficiency properties of energy-only markets are therefore undermined by real world concerns.

§ Capacity Payments: Britain and Spain both defined capacity payments in terms of a price available to all capacity. The market in Britain used a complicated formula to calculate the probability of an outage and hence the likelihood that capacity would have a high value. However, these rules were subject to manipulation (and ceased to reflect real world conditions after some time). The Spanish rules were much simpler, but lacked the coherent rationale needed to offer stability in the long-run. Both systems ensure that the price paid to capacity falls when the volume of available capacity increases and vice versa, in order to dampen the effect of over- or under-estimating the price.

§ Capacity Obligations: A key feature of systems in the US, these obligations make up for the effect of energy market price caps, which deny generators the chance to earn peak prices needed to remunerate capacity. The schemes set a required quantity and let the price vary in the market, as necessary to call forth the required quantity of investment. Such schemes have evolved over the years, to cope with wholesale and retail competition, and are continuing to evolve. A key requirement seems to be the need to avoid major fluctuations in the market price of capacity, which requires a system of monitoring capacity some time in advance of the actual peak.

§ Hybrid Schemes: Just as capacity payments adjust in the light of market conditions, so the New England capacity obligation is now being defined in a way that reflects the observed price of capacity. Both types of scheme therefore converge on a system where the reward for capacity is defined by a “downward-sloping demand curve”. This aspect of the scheme is difficult to define objectively, but dampens the effect of errors in estimating capacity requirements. The scheme also allows the achieved level of security of supply to adapt to differences in the cost of meeting it, which might be important in the SEE market, where the cost of building new capacity is not known with much certainty.

This summary of the options indicates that each scheme has some advantages and disadvantages. Our appraisal indicates that PPAs and hybrid capacity obligation schemes are prefereable in the conditions of South East Europe. However, even within the context of the SEE regional electricity market, conditions in each member country differ so much that no single scheme is likely to be optimal in all countries. Attempts to impose a common scheme immediately are therefore unlikely to provoke agreement, or indeed to be efficient. What is needed instead is a process whereby the members of the SEE regional electricity market can set out the parameters of a transitional programme that ensures SEE countries are working towards an ever more integrated system, without causing problems for each other in the interim. Section 6 attempts to set out such a process, with the aim of implementing a recommended system in due course.

6. Recommendations for SEEREM

Certain preconditions must be in place before any capacity scheme can begin to take effect and before investors will have confidence in market-based schemes. The following sections explain these preconditions and also set out priorities for agreeing the terms of an eventual capacity scheme.

6.1. Offering a Reasonable Prospect of Cost Recovery

No capacity scheme will be effective in encouraging investment by the private sector if investors doubt whether the costs they incur under the scheme will be reimbursed. Overcoming this hurdle is a problem in two cases (which may occur simultaneously in some countries):

- If regulated retail tariffs produce insufficient revenues to cover the existing costs and liabilities of the electricity industry, or cannot be raised by enough to cover the new costs and liabilities caused by new investment in generation; and
- If the prospect of retail competition threatens to turn long-term investments and contracts into "stranded costs" which cannot be recovered because of the impact of competition.

A precondition for successfully attracting private sector investment is therefore the ability to convince investors that they can recover their costs, whether via a market or via a long-term PPA. SEE member states will need to demonstrate the reality of this precondition, by setting tariffs that cover costs and showing a willingness and ability to raise tariffs when costs rise.

Some countries will have passed this hurdle already, but until all countries in the SEE region have the same reputation, investors' decisions will be biased towards those where cost recovery is more secure, regardless of the relative need for capacity.

6.2. Selective Use of Long-Term PPAs

In some SEE countries, market institutions will take time to develop and real competition among creditworthy buyers and sellers will take even longer. In these conditions, investors will not have confidence in their ability to secure reasonable prices through short term trading and they will require the financial security offered by a long-term PPA. (Of course, to be able to offer long-term financial security, the buyer of power under the PPA must have a reasonable prospect of cost recovery.) It is therefore likely that some SEE countries will need to rely on PPAs for some years to come.

If relying on PPAs, each SEE country will need to set up a system to establish capacity requirements, to impose an obligation on a single buyer to procure the necessary capacity, and to anticipate future market developments. The system for establishing capacity requirements and the definition of the obligation both need to be coordinated throughout the SEE region, as discussed below. Each SEE country will then need to set out the procedure for dealing with stranded costs, if wholesale or retail competition should enter the market before the PPA expires.

This "procedure" for handling stranded costs may be little more than a legal duty placed on the energy regulator to allow the single buyer the right to recover prudently incurred costs, in accordance with EU directives on the internal electricity market and state aids. To support the application of the rule, however, it may be necessary to anticipate future challenges by ensuring that the PPA is "prudently incurred", by documenting fully how the buyer identified a need for capacity and how the selection process minimised the associated costs. The investors that wins such a competition to build capacity will be able to offer lower prices if such matters have already been clarified.

6.3. Capacity Obligations

Fundamentally, it will be important for the members of the SEE regional electricity market to work towards a harmonised (though not necessary common) system of capacity obligations. The process of moving towards a harmonised system includes the following steps.

1. **A regional agreement allocates capacity obligations to each SEE country**

2. **Each SEE country allocates its obligation among one or more parties (TSOs or retailers), if necessary by zone**

A clear definition of capacity obligations will in any case be necessary to define the required volume of PPAs during the transitional phase. However, the obligation should be made explicit, under a regional agreement, in order to ensure that no single SEE country will be able to "lean" on the capacity of its neighbours, thereby bringing down the electrical system. (This problem is not specific to the SEE region. It forms the basis for the creation of similar rules in the US.)

The regional agreement should set a basic target for capacity (e.g. 100% of a peak demand forecast plus a reserve margin of X%). The rules will have to define the lead time of this obligation, i.e. the year to which it applies. [30]

The rules should allow any SEE country to move from a fixed obligation towards a system with a "downward sloping demand curve". The flexibility offered by this scheme has advantages in SEE conditions. Some countries may find that the cost of meeting a common capacity requirement is prohibitive and that customers would prefer to suffer a lower level of security in return for lower tariffs. Such a choice would be an efficient response to information that building capacity is expensive.

Incorporating the demand-curve idea provides an automatic and relatively transparent basis for this kind of adjustment. Box 1 sets out some proposed parameters to define this demand curve. In practice, it is difficult to apply in any but a fully blown system of capacity obligations, but should not be ruled out even at this stage by any regional agreement.

---

[30] The lead time is the period of time between the date upon which mutual commitments are formally agreed, and the date upon which performance provisions first apply (i.e. the date by which a generator must have capacity available). The length of the lead time may vary from nothing, to a period of months or years. Longer lead times can allow new entrants to compete for capacity payments. However, the total quantity of capacity requirement is less certain when the lead time is longer. The Greek proposal anticipates a one year lead time.

Note that capacity obligations should in principle be defined by zone, when there are significant transmission constraints that prevent capacity in one location from meeting demand in another. It is not clear whether any SEE country needs to divide its capacity obligation by zone, but the agreement should allow for this possibility.

Box 1: Parameters of a Demand Curve for Capacity

The parameters of this demand curve would be:

C^* = annualised cost of peaking capacity

Q^* = Basic target level of capacity = Forecast peak demand *plus* X%

= Level of capacity at which the value of capacity is C^*

Q_{max} = Level of capacity at which value of capacity is zero
= Q^* *plus* Y%

Q_{min} = Level of capacity at which value of capacity reaches a maximum of ($2 \times C^*$)
= Q^* *minus* Z%

The values of Y and Z would be set so as to define the widest acceptable level of variation in the capacity margin. (For reference, the values proposed for New England by Cramton and Stoft are X=18%, Y=11%, Z=4%.)

3. Set up a scheme in each zone for measurement and registration of capacity holdings

The next priority should be a system for measuring and registering the availability of capacity from each generator, based on its expected contribution to meeting peak demand. This will require some standard rules for the measurement of capacity at hydro plants and other special cases. In principle, each SEE country can establish its own rules for measuring capacity.

These rules should set out the extent to which capacity in neighbouring countries is accepted in fulfilment of an obligation. To begin with, SEE countries may be reluctant to accept capacity from any of its neighbours, but bilateral agreements should enable some trade in capacity, which is likely to lower total costs. These bilateral agreements would set out:

- Which capacity in country A is accepted as fulfilling obligations in country B;
- What transmission capacity between countries A and B traders must hold (and for how long), if they wish to assign generation capacity in country A to an obligation in country B;
- What registration procedures would be necessary to ensure that the same generation capacity cannot be used to offset obligations in country A and country B (unless their peak demands never coincide).

In addition, it will be necessary to spell out who will be the owner of any capacity still covered by a PPA. It may be that such capacity belongs to the TSO, or to some other agency, but it will require compensation for its holding under any generalised scheme.

4. Define the terms of the capacity contract that is accepted against the obligation

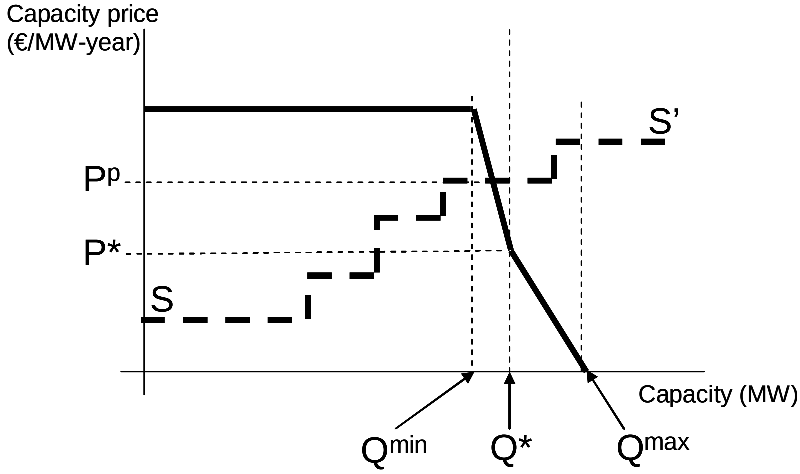
If retailers want to show that they have capacity to fulfil an obligation, they must be able to demonstrate that the generator is subject to some obligation to make capacity available when it is needed. In SEE countries that have (or are proposing to adopt) a mandatory spot market (or “gross pool”), this contract would specify that generators have to offer their full capacity to the spot market (or else to suffer the penalty for being short of capacity). In SEE countries that allow traders to notify bilateral contracts to the settlement system, the contract would oblige generators to offer uncommitted capacity to the balancing mechanism.

Ideally, the contract would also say that generators must offer a rebate to the holders of their capacity if the price they receive from the spot market or balancing mechanism rises above a certain level. However, many capacity schemes operate without this clause and it may be rendered unnecessary by an explicit or implicit cap on energy prices.

5. Develop a method for TSOs to buy and pay for capacity to make up shortfalls

To ensure the capacity market works efficiently, the TSO in each country should establish a market to make up any short-term gaps between the capacity holdings and capacity obligations. The TSO would accept offers from generators whose capacity is not already committed and would compare those offers with the “downward-sloping demand curve”. Figure 6.1 shows a set of capacity offers (dashed line SS’), in which the first “step” represents capacity already held by retailers. Demand is represented by the downward-sloping demand curve defined by the prices and volumes of capacity (solid line). At the point where supply and demand match, the TSO sets both the price and volume of capacity requirements.

Figure 6.1
Capacity Market



If capacity obligations had been allocated among the retailers, rather than retained by the TSO or a central procurement agency, the TSO would then allocate the capacity it had bought to retailers and would charge them the associated price.

6. Pass through costs to customers

The final stage in the process returns to the first pre-condition, and requires that retailers be allowed to pass through the costs of buying capacity to retail customers.

The final advantage of the downward-sloping demand curve is that the price of capacity rises slowly, as excess capacity is gradually eliminated. Thus, if actual capacity is above Qmax, the price of capacity will be zero. Only when it starts to fall below that level does the price rise above zero, and then only gradually. If retailers hold long-term PPAs that count towards their capacity obligation, their own costs may not rise when the capacity price rises. However, year-by-year, retailers' contribution to capacity costs will rise slowly and will provide an objective reason for increasing retail tariffs.

Appendix B provides a more detailed description of the design of a capacity obligations scheme.

6.4. Conclusion

The SEE region contains countries that face a wide range of conditions. All will require investment in capacity in the mid-term, but they are at differing points in the evolution of their electricity markets. Some markets are, or are about to come, under the requirements of EU directives, whilst others are many years from EU membership. Some have a good track record on cost recovery, whilst others are struggling to put their electricity sectors on a sound financial footing. These differences mean both that a common system is unlikely to be optimal in all countries, and that there are dangers in relying on any common system, in case some countries are unable to fulfil the necessary preconditions.

The recommendations set out foresee a transition from the use of PPAs towards a more market-based system, based on capacity obligations. Work on some elements of the scheme – particularly the definition of capacity obligations – can begin at once. Some elements will require bilateral agreements between SEE member states – such as the recognition of capacity in neighbouring states. However, proposal set out here is intended to provide a common agenda on which all countries in the SEE regional electricity market can focus their attention.

Appendix A. Market Paradigm

Most electricity markets are intended to set prices in relation to the variable operating costs of generation (i.e. unit costs of energy production), but such prices would not provide adequate remuneration for investment in thermal generation capacity. Some mechanism is required to move prices above variable operating costs for at least some of the time.

We describe below a well-known market paradigm that explains how prices in a competitive electricity market incentivise investment in a socially optimal portfolio of generation. As will be seen, this paradigm relies heavily on the ability of short-term electricity market prices to soar to very high levels during a shortage, in order to remunerate the investment in generation capacity needed to meet peak demand. There are several reasons why policy makers might be concerned that this market paradigm will not function well and why, as we will explore in this paper, some additional scheme may be required in order to incentivise adequate capacity investment.

A.1. Generation Cost Conditions

The paradigm begins with the recognition that electricity is generated by a variety of technologies, each of which has different cost characteristics.[31] For clarity of exposition, the paradigm is normally presented in terms of three technologies as follows:

Table A.1
Cost Characteristics of Typical Production Technologies

Type	Annual Fixed Cost	Variable Cost per Unit
"Baseload"	High	Low
"Mid-Merit"	Mid	Mid
"Peaking"	Low	High

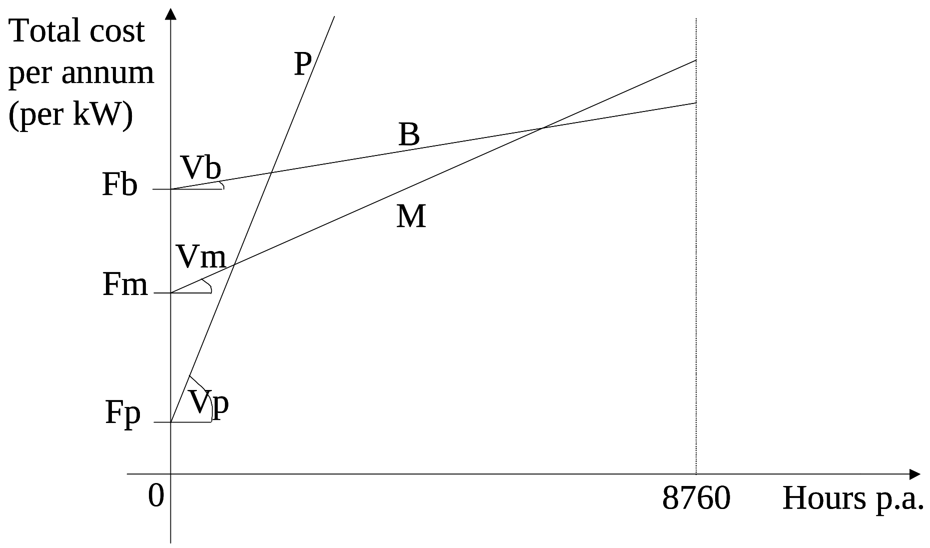
In practice, the nature of these technologies changes from time to time. In 1990, baseload generators included both "must-run" plant (run-of-river hydro and nuclear) and coal-fired generation; mid-merit plant meant oil-fired generation. By 2000, however, gas-fired generation had largely supplanted the role of coal as a baseload technology and coal-fired generation had been pushed into mid-merit. Recent rises in gas prices and the change in market arrangements have reversed this trend to some extent. These changes show the risks associated with long-term investment in generation and the associated pressures for risk management (for example, diversification of fuel sources). The paradigm takes costs as given and users must therefore make separate allowance for risk.

---

[31] One may ask why cost conditions should fall into this neat pattern and the answer is "by elimination". Any plant for which both costs were high, or for which one cost was high and one middling, would be uneconomic compared with at least one other technology. As a result, it would be ignored or devalued, until either its fixed or variable costs fell sufficiently to make it competitive against the technologies shown, in which case it would replace one of them. In principle, costs need not be only "high", "middling" or "low", but can take any value. As a result, it is possible for many technologies to be economic. However, in practice, it is rare for more than three technologies to achieve least-cost status at any one time. Other technologies may be left over from previous eras but must compete with those that are currently least-cost by revaluing their capacity until their fixed cost is low enough to be competitive against hew technologies. If achieving this position requires a negative value for capacity, the plant should close.

Figure A.1 shows the total annual costs (per kW of capacity) for the three basic types of generation: baseload (b), mid-merit (m) and peaking (p). Each type of plant is represented by an annual fixed cost per kW (F) and a variable cost per kWh (V). Total annual costs therefore vary according to the number of hours that capacity runs in a year, up to the maximum of 8760 hours (= 24 x 365). For example, a kW of baseload capacity has a relatively high annual fixed cost of **Fb**, which the plant incurs even if the plant does not run for any hours during the year. Hence, at 0 hours (on the left hand side of the figure), the annual cost per kW for baseload capacity is Fb. If the plant runs, its costs per kW increase, at a rate equal to **Vb**, the unit cost of output from baseload capacity. The line marked **B** therefore represents the relationship between annual hours of operation and total annual costs for baseload plant. (This report will refer to "hours" even though electricity markets may use half-hours or even shorter periods for settlement purposes. The analysis applies for different settlement periods, but the arithmetic is complicated by time period conversion factors.)

Figure A.1
Annual Costs per kW for Different Types of Generation



Mid-merit plant has a slightly lower annual fixed cost per kW, **Fm**, but a slightly higher variable cost per kWh, **Vm**. The cost function of mid-merit plant is given by the line marked **M**. At zero hours of operation, the cost of mid-merit plant is lower than the cost of baseload plant, but its costs rise rapidly if it runs. Mid-merit plant is more expensive than baseload plant if running for most of the year (ie, line **M** rises above line **B** on the right hand side of the graph).

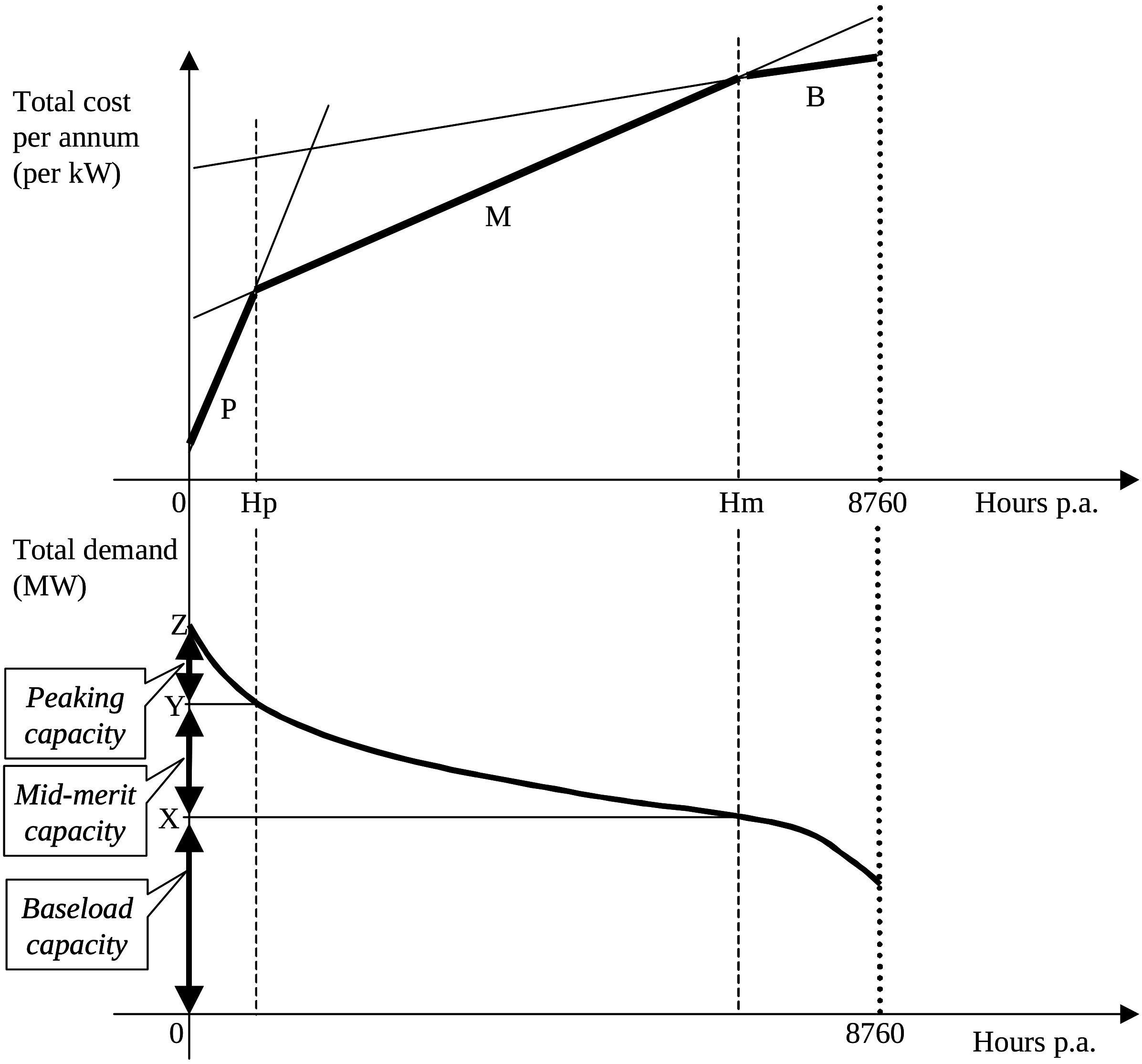
Similar, for peaking plant, the annual fixed cost per kW is very low, at **Fp**. However, its variable cost per kWh, **Vp**, is very high, so that the total annual costs of peaking plant are very high, if it runs for a lot of hours in the year.

Given these three technologies with different cost conditions, it is possible to calculate an efficient pattern of investment in each one, sufficient to meet peak demand with the least total cost.

A.2. Efficient Least-Cost Investment

In planning terms, it is possible to identify an efficient portfolio of investment, by comparing the cost conditions in Figure A.1 with the demand conditions shown in Figure A.2. Here, the top half of the figure shows the same three cost functions, **B**, **M** and **P**, as in Figure A.1. The darker line indicates the least-cost technology at different levels of operation over the year. If the plant is expected to run for only a few hours (**Hp**), then peaking plant is the cheapest form of generation, because its low fixed costs outweigh its high running costs. For plant expected to run between **Hp** and **Hm** hours, mid-merit plant has the lowest costs overall. For any plant expected to run more than **Hm** hours per year, baseload technology is the cheapest, as its low variable costs more than compensate for its high fixed cost.

Figure A.2
Efficient Capacity Planning Model



The lower half of the figure shows a "load duration curve", a conventional presentation device in the electricity industry that shows the level of demand in particular half-hours over a year. On the left hand side is the hour with highest demand; on the right hand side is the hour with lowest demand. Other hours are arranged between them in descending order of demand. Reading horizontally, one can see the number of hours ("duration") in which demand is equal to or higher than any particular level. (For planning purposes, the measure is a forecast, based on historical data.)

The figure shows the boundaries **Hp** and **Hm** transposed onto the load duration curve. The interpretation of this graph is that a least-cost portfolio of generation would include baseload capacity of OX, mid-merit capacity of XY and peaking capacity of YZ. Then if, in any hour, generation capacity operates in least-cost order *with respect to its variable costs*, each plant will run for a number of hours per year at which it represents the least-cost technology. Baseload capacity will run for long periods (greater than **Hm**), ie nearly all of the time. Mid-merit plant will run for up to **Hm** hours, but will be restricted to hours when demand is higher than OX. Peaking plant will only run when demand exceeds OY, and will not run for more than **Hp** hours.

A.3. Energy Revenues, Capacity Payments and Security Standards

Even under monopoly systems, central planners noted the implications of this paradigm for tariffs and revenues. They assumed that the efficient (wholesale and retail) price per kWh of energy would be the "system marginal cost", ie the marginal cost of meeting an extra unit of demand. In hours to the right of **Hm**, that cost is represented by **Vb**, the variable cost of additional output from baseload capacity. In hours between **Hp** and **Hm**, system marginal cost is **Vm**, the variable cost of mid-merit capacity. For hours below **Hp**, the system marginal cost would be **Vp**, the variable cost of peaking capacity, but for the following complications.

First, the capacity planners noted that, if prices or tariffs only reflected variable costs, they would not recover the whole costs of the system – which would prevent efficient investment from taking place. Further examination of the cost structure set out above proved that the "missing" amount of revenue was equal to **Fp**, the fixed annual cost of a peaking generator, multiplied by the total amount of capacity on the system. This observation led to the design of various "capacity payments" to augment energy charges.

Second, planners noted that building generation capacity sufficient to meet demand of OZ was inefficient, since consumers were not willing to pay as much as **Fp** for an extra unit of energy in the single hour of peak demand. They usually adopted a security standard, for instance a policy of anticipating 10 or 20 hours per annum of "lost load". Figure A.3 is merely a close-up of the section of peak demand section in Figure A.2, but shows the rational for such standards.

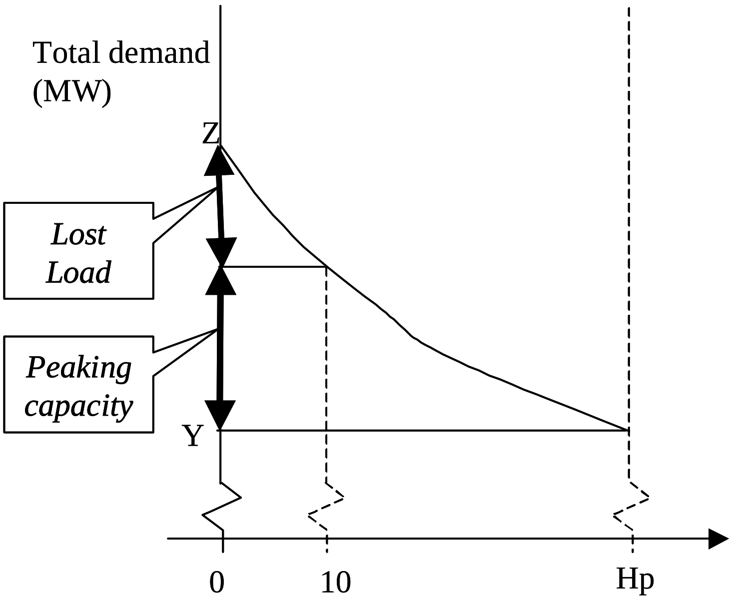
In this case, the planners have decided to build total capacity that is sufficient to meet total peak demand in all but 10 hours. In those 10 hours, demand would rise higher than the level of total capacity, so some demand must be cut-off ("lost") by the system operator. In – principle, the efficient level of lost load depends on a trade-off between capacity costs (**Fp**) and the "value of lost load" (VOLL).

Suppose that the fixed cost of building and maintaining 1MW of peaking capacity (**Fp**) is equal to £20 000 per year. Suppose consumers are willing to pay £2 000/MWh (VOLL) to maintain supplies at times of peak demand. (Assume also that variable costs, **Vp**, are relatively small, compared to £2000/MWh and can be ignored here.) The planners have a choice:

1. Build one more MW of peaking capacity at a cost of £20000 per year; or
2. Cut off 1 MW of load for 10 hours per year.

The second option means an extra 10 MWh of lost load per year, at a value of £2000/MWh, giving a £20 000 per year total cost of the lost load. In this example, the customers are indifferent between spending £20 000 on the new capacity, or losing £20 000 from the lost load. If there were less peaking capacity, there would be more hours of lost load, and option 1 ("Build") would be cheaper than option 2 ("lose load"), so there would be a signal to invest. If there are less than 10 hours of lost load per year (on average, taking several years together), the revenues for capacity will not be sufficient to recover the fixed costs, because the customers are not prepared to pay more than £2000/MWh. This will provide a signal that it may be efficient to close plant. A similar set of comparisons between the costs of peaking, mid-merit and baseload stations will show what type of plant should be built, depending on the number of hours it is expected to run. In practice, the answer is usually a peaking plant or a baseload plant: mid-merit capacity is nearly always old plant displaced from a baseload role by newer, more efficient technologies.

Figure A.3
Peak Demand Conditions



The choice of security standard (i.e. the desired average annual number of hours of lost load) therefore determines the level of investment in capacity in a planned system. This approach is not directly relevant to a market system, but it can be adapted to show what an electricity market would look like, if it offered similar incentives for efficient, least-cost investment.

A.4. Extending the Market to Cover Market Conditions

To convert the analysis above into a market paradigm, it is necessary to recognise that demand-response – in the form of load-shedding – is an alternative method of balancing supply and demand, to be placed alongside the generation technologies shown in Table A.1. For simplicity, we will examine only the possibility of losing load at peak times and will refer to a single value of lost load, meaning that value which applies at peak times, when consumers are being forcibly interrupted. Table A.1 can then be extended with a "fourth technology", losing load, which has a zero annual fixed cost, but a very high variable cost per kWh of lost load, as in Table A.2.

Table A.2
Extend Table of Cost Characteristics

Type	Annual Fixed Cost	Variable Cost per Unit
"Baseload"	High	Low
"Mid-Merit"	Mid	Mid
"Peaking"	Low	High
"Lost Load"	*zero*	*VOLL*

This information can be added quite simply to the diagram of cost conditions and capacity planning, by incorporating an additional line, **L**, to represent the possibility of losing load.

This line creates another cross-over point, Hl, below which losing load is cheaper than building and using peaking capacity. The efficient decision is therefore only to build capacity up to the level now marked as "C" and to shed load when demand exceeds that level. The volume of lost load is shown in Figure A.4 on page 49 as a shaded triangle in the top left-hand corner of the load duration curve.

A.5. Implications for Electricity Market Pricing

The implications of this adaptation become apparent when applied to a competitive electricity market, given two key assumptions about the way in which competitive markets work:[32]

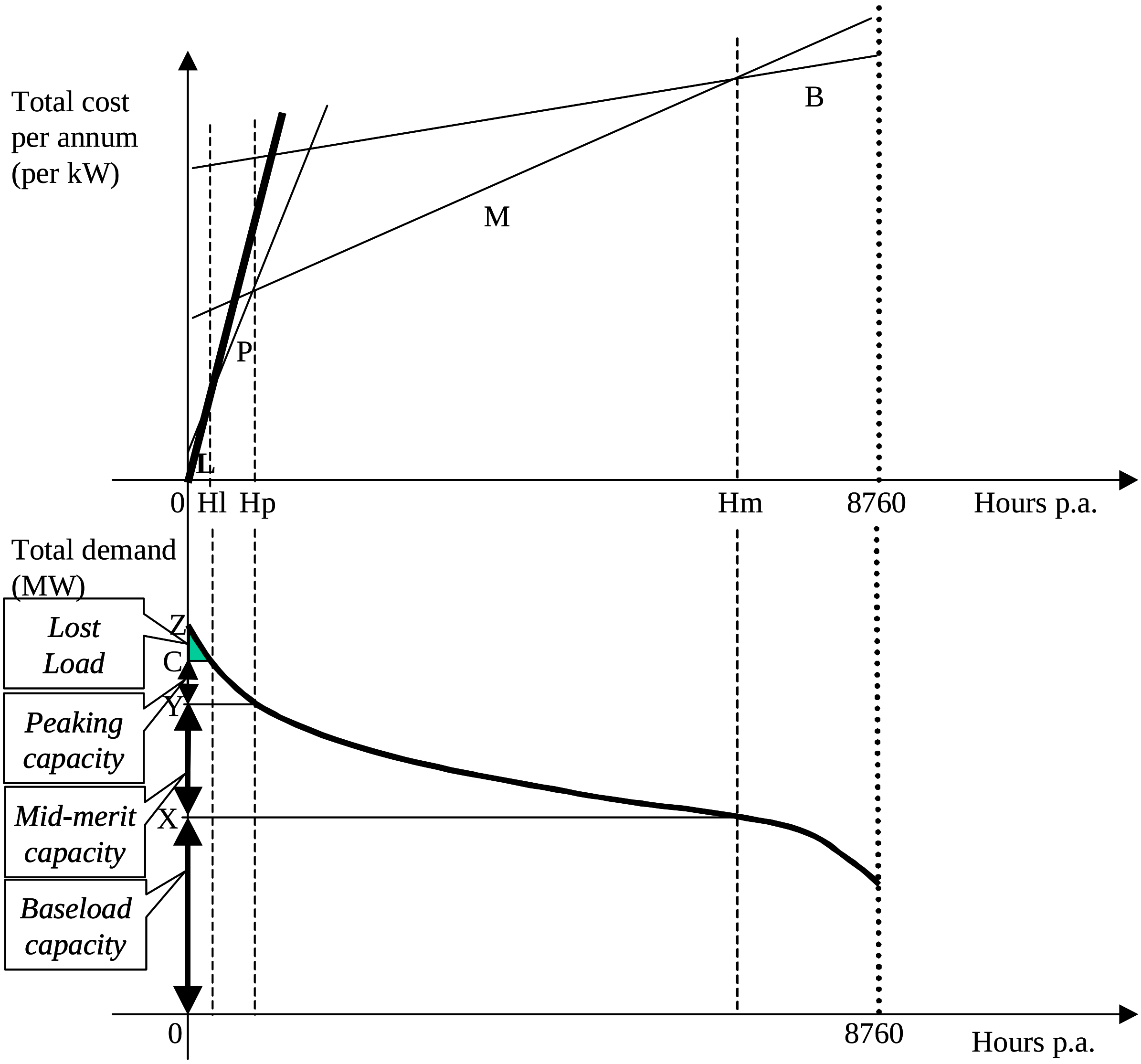
1. **Producers will select the least-cost combination of output.** Given that fixed costs are unavoidable from hour to hour, producers will minimise variable costs by running only baseload plant when demand is low, calling on mid-merit plant only when demand rises above OX and using peaking plant only in the rare hours when demand exceeds OY. Producers will continue to serve demand as long as they have capacity available and load-shedding will only be necessary if demand would otherwise exceed OC.

2. **Market prices will settle at the marginal cost of the most expensive producer chosen or at VOLL.** Competition normally drives down prices to marginal costs and the electricity sector is no exception to this rule, although the marginal costs for the market

[32] The decision to replace monopolies with markets rests on the belief that these conditions apply in markets or apply at least as much in markets as in monopolies, or do not apply at least as much as under a monopoly, it is open to question whether markets offer any advantage over monopolies in the electricity sector.

(or "system") will be determined by the plant type with the highest variable cost that is called upon in the least-cost combination.

Figure A.4
Market Determination of Capacity Investment



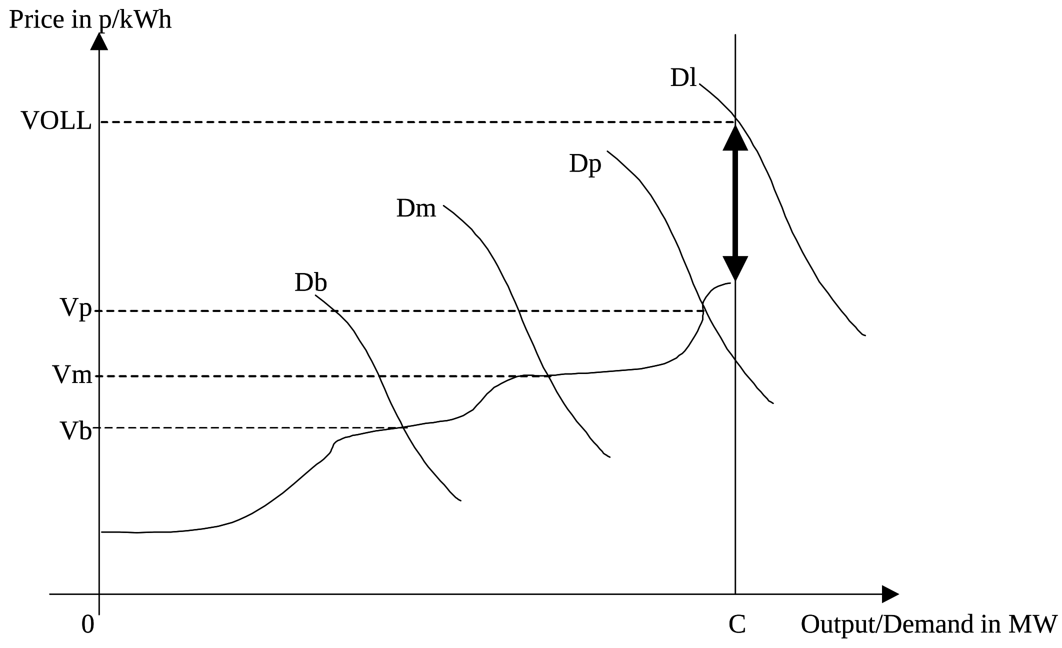
As before, this means that the price of energy will equal the "system marginal cost", but in the hours of lost load there is a new definition of this term. Instead of peak prices being set equal to the variable cost of peaking plant, **Vp**, with some capacity payment required to recover total costs, prices are now set equal to VOLL – the variable cost of shedding load – in the few hours when demand exceeds capacity.

The revenue from sales at VOLL in these hours substitutes for the capacity payment described in section A.3 and, in an efficiently built system, will equal the capacity cost of a peak (as per the analysis in that section). If there are a lot of hours of lost load, these revenues (annual hours of lost load x value per kWh of lost load) will exceed the cost of adding capacity (annual cost per kW) and will encourage investment. If there are few hours of lost load, there is excess capacity and it may be efficient to close plant.

In an electricity market, therefore, the incentive to invest depends crucially upon the market price that applies when load is being lost, and the expected number of hours in which load will be lost over the life of the investment. Box A-1 provides an equivalent analysis in terms of conventional supply and demand diagrams.

Box A-1
Supply and Demand Analysis

In the diagram below, demand can take one of four (typical levels), labelled by reference to the plant type required to operate: baseload (Db); mid-merit (Dm); peaking (Db); or exceeding capacity (Dl). In each case, the market price is given by the variable cost of the marginal plant (Vb, Vm, Vp) or VOLL in the case where load is being shed. The incentive to invest in generation capacity depends on the frequency with which these (typical) demand conditions arise and, in particular, the frequency with which demand reaches Dl, such that load is lost, the price rises to VOLL and generation capacity receives the additional revenue needed to cover fixed costs, as marked by the double-headed arrow.



Appendix B. Capacity Obligation – Detailed Terms

B.1. Product Definition

The first step in designing a capacity scheme, whether it be a fixed payment CPM (capacity payment mechanism) or a fixed quantity CPM, is to define the product which is being provided and paid for. While the product is generally referred to as "capacity", there is no universal definition of what is meant by capacity. It is necessary to define in detail what the generator's obligations are in providing capacity, how compliance with those obligations will be monitored, and what the "buyer's" obligations are, including the details of the payment scheme. Essentially, the definition of the capacity product is a description of the terms and conditions of the capacity contract and the capacity product is fully described by the contract. (NB – the capacity contract need not be an explicit bilateral contract, but may be established as part of the market rules). For purposes of this consultation, only the principal features of the product design will be reviewed.

B.1.1. Term

The term (tenor or length) of the contract is a significant factor. The term sets forth the period for which the mutual commitments will apply.

When capacity payments are embedded as part of the market rules, the term is usually indefinite as the payment scheme will apply until those rules are changed. Changes in market rules will require the approval of the regulators and in some cases may require legislation or ministerial approval. The rules have no fixed term as the prerogative of regulators or legislators to change rules cannot be given away. However, even with an indefinite term, the payment may be fixed for a certain amount of time. For example, the obligations of a generator and the payment methodology may be set by market rules and will apply until those rules are changed, but the rules may call for a central authority to set the capacity payment and fix it for six months, a year, or a number of years. In this case, the term of the payment is definite even though the contact term is not.

Alternatively, it is possible that the market rules may call for a central authority and each generator to enter an explicit capacity contract that would set forth mutual obligations. In that case, a key element of the contract would be the term, which could also vary from a period of months to a period of years. Longer terms, either with respect to the duration for which a payment is fixed or with respect to an explicit contract, can be an important factor in the risk that investors assign to projections of CPM revenues.

B.1.2. Lead Time

The lead time is the period of time between the date upon which mutual commitments are formally agreed, and the date upon which performance provisions first apply (i.e. the date by which a generator must have capacity available). The length of the lead time may vary from nothing, to a period of months or years. Longer lead times can allow new entrants to compete for capacity payments, however the total quantity of capacity requirement is less certain when the lead time is longer.

B.1.3. Performance Provisions

Performance provisions are the definition of the obligations of the generator, and can also determine the basis for payments and penalties. The performance provisions define what a generator must deliver to receive payments and what a generator must deliver to avoid penalties. Since a main objective of capacity is to meet a supply adequacy standard measured in LOLE, the performance requirements are geared to factors which enable LOLE to be reduced by the development and maintenance of generation capacity. Key aspects of these provisions are as follows:

§ *Demonstrated Maximum Net Capability (DMNC).* A generator will generally receive a capacity payment based on its DMNC. This measure of capacity is established by periodic tests that require a generator to demonstrate that it can produce power at a given level for a given duration at specified ambient conditions. Typically, a generator may be required to demonstrate every six months the level to which it can produce over several hours. The achieved level of output is then adjusted to the equivalent production level in specified ambient conditions (e.g. a defined winter temperature). The old England & Wales pool adopted an alternate system, in which generators' availability was set equal to their declared level (subject to a cap equal to the registered capacity for which they paid transmission tariffs), until they plant suffered a forced outage, after which the cap was set equal to the maximum level of output achieved since the outage, This method avoided the need for periodic tests (although the rules allowed for such tests in the case that a generator's declaration of availability was not believed). Under either scheme, a generator would be eligible to receive capacity payments for no more than the MW level of generating capacity that can be demonstrated. In performing LOLE analysis, the system planner would also represent each generator as capable of supplying output up to its DMNC. The most fundamental performance characteristic of the capacity obligation placed on generators is therefore the DMNC.

§ *Primary energy source availability.* The ability of a generator to reduce LOLE will depend not only on its DMNC, but also on the availability of its primary energy source (fuel). Generators with virtually unlimited primary energy sources contribute to reducing LOLE based on DMNC and mechanical availability. However, generators with limits on fuel availability (such as hydro plants and wind turbines) can reduce LOLE only when fuel is also expected to be available. For such generators, the impact of fuel availability must also be considered when determining the value of capacity. One method of doing so is to compute the equivalent capacity that an energy-limited source contributes to reducing LOLE, by using modeling that incorporates the energy source limit and determines how much capacity with a given expected pattern of primary energy source availability would be required to provide as much benefit in reducing LOLE as one MW of capacity without any primary energy source limitations. Ideally, such an analysis would also account for any correlation between primary energy source availability and system load. For example, hydro sources that have significant storage are often counted as fully equivalent to non-energy limited resources, even though they are actually energy limited, because production is controllable and can be directed toward periods when the capacity is needed. Uncontrollable primary energy sources such as wind often count as less than fully equivalent because they may be unavailable when needed. Given data concerning the expected pattern of production, variance in the expected pattern and correlation with factors affecting system load, stochastic models can be used to determine capacity equivalence. In cases where there are energy limits resulting from limitations in the primary energy source (or limits of any kind in hours of use that are unrelated to physical availability), the DMNC is usually adjusted to reflect the equivalent unlimited capacity.

§ *Mechanical Availability.* Mechanical availability also affects the degree to which generating capacity can be relied on to reduce LOLE. Mechanical availability is represented by a forced outage rate and maintenance requirements. All else equal, capacity with higher forced outage expectations or greater maintenance requirements contributes less per unit of DMNC to reducing LOLE, and is usually eligible for a lesser payment than capacity which is more available. CPMs usually require that performance be tracked and verified with respect to scheduled maintenance and forced outage, and also that this performance in some way affect the payment or trigger a penalty. As LOLE is modeled assuming a coordinated maintenance schedule, it is also common that participation in a CPM involves a requirement for coordinated maintenance, meaning that generators should inform the system operator of their maintenance plans and should consider requests by the system operator to move maintenance outages.

§ *Deliverability.* In order to reduce LOLE, the output of the plant must be deliverable under peak conditions. However, some plants may be unable to operate at full output at peak times, due to network constraints. To recognize this possibility, capacity requirements or payments could be defined on a locational basis – the standard being to define them for broad zones, rather than individual nodes. Alternatively, to qualify for a capacity payment, the plant could be required to meet a defined deliverability standard. Particularly if there are only shallow interconnection requirements, a deliverability requirement may be necessary to ensure that a plant which does not enhance reliability (i.e., reduce LOLE) does not receive a capacity payment. In other words, it may be advisable to check how much of a plant's potential output is deliverable to all load under peak conditions. In addition, since LOLE is modeled assuming that units will be available at times of peak system stress (unless it is affected by a forced outage), it is common that participation in a CPM requires generators to offer energy to the market at times of TSO-declared peak system stress, unless on forced outage.

B.2. Payment Structures

B.2.1. Capacity vs. Availability Payments

A capacity scheme can be designed upon the assumption that energy (and ancillary service) markets are the primary tools to encourage short-term availability. Alternatively, the CPM payment structure can itself be designed to be the primary tool to encourage short-term availability. There are several options:

§ *Unforced Capacity Payment:* Under an unforced capacity payment methodology, each unit would be paid a fixed annual capacity payment in equal monthly installments based on its unforced capacity (capacity times one minus the forced outage rate). The payment would serve as an incentive to construct and maintain sufficient capacity. The payment would be affected by actual availability in that lower or higher availability would reduce or increase the quantity of unforced capacity that a unit would be eligible to sell. The unit would earn the payment by establishing DMNC through tests and establishing forced outage rate history. An open issue is how to measure forced outage for units that operate infrequently.

§ *Availability Payment:* An alternative approach is that each unit could be paid based on half-hourly availability. To earn the payment in a half-hour a unit would need to be available for that half-hour. Available units would be paid the capacity payment on annual basis divided by: the number of half hours in the year divided by one minus a target on availability factor, say 10%. Hence a unit available 90% of the target would receive the full capacity payment. Higher or lower availability would result in higher or lower revenue. While units would still be required to demonstrate DMNC, payment would be based or declared half hourly availability up to this DMNC. This methodology is appealing if the energy market caps are so tight that they do not sufficiently encourage short-term availability. This is methodology is inefficient, however, to the extent it might require certain plant that could be shutdown on weekends or for period of weeks in the summer to remain available just to receive the payment.

§ *Shaped Availability Payment:* A third alternative is a variant of the second. Payments would be based on declared availability by half-hour. The aggregate payment level available to a generator that was available for 90% of all half-hourly periods would be equal to the full annual capacity payment level. However payment would be shaped to reflect seasonal and diurnal loss of load probabilities. Payments would be extremely low over low-load night periods, highest in the winter day periods and relatively low in the summer day times. Seasonal variations would be mitigated by the impact of scheduled maintenance on LOLP. This methodology would still require that generators be available in each half hour to earn a capacity payment. However the shaping of such payments mitigates somewhat the potential inefficiency of requiring generators to be available. A peaking generator with material stand-by costs could make a decision to be unavailable in low-load periods, for example, and the economic cost of it making this decision would be less than under the second option.

The latter two payment structures do require that availability be declared on a half-hourly basis. A unit with say a six-hour start time from cold conditions would not physically be available at short notice if not chosen by the central unit commitment. However, for purpose of the capacity payment, if not committed, a unit would nevertheless be deemed available if it was offered into the unit commitment.

B.2.2. Contract for Difference Payment Structures

It is very difficult to estimate the net operating margin that a peaking unit would expect. Understatement would result in overpayment by customers; overstatement would result in an insufficient incentive for entry. Both are significant problems.

A payment structure that mitigates this problem is a one-way Contract for Difference (CFD) type of structure. Under a one-way CFD, the capacity payment to a generator would be offset by the imputed margins the generator could have earned in the energy market over and above a predefined strike price. Specifically, for each half-hour period, a generator with such a one-way CFD would potentially be credited with a CFD offset, and the sum of these half-hourly CFD offsets would be subtracted from the capacity payment made by the central authority to that generator at the end of each month. The CFD offset in each half-hour period would equal the maximum of:

§ Zero; and

§ The energy market price in that half-hour period, minus a predetermined CFD strike price, all multiplied by a MW CFD quantity specific to that generator.

The strike price (and other CFD parameters, if any) would be common across all CFDs, so that from the central authority's perspective, all capacity is considered equal – and is priced accordingly. For example the strike price might be based on the variable cost of a peaking unit. The monthly capacity payment would be offset by the sum over all half-hours of the excess of market prices over this CFD strike price.

The use of such a CFD payment structure could be beneficial because it could lower overall risk associated with the CPM. The net capacity payment would hedge both customers and generators against market volatility. Instead of estimates of expected profits which may be wrong, a more realistic reflection of profits would apply. If energy prices were in excess of peaking operation costs for a significant number of periods, the net capacity price paid would decline while energy prices and profits would rise. If, on the other hand, prices provided few profit opportunities for peaking generators, net capacity prices would stay at the full level, but energy profits would be low. Under this structure, neither customer nor generator would bear the full risk of an inaccurate projection of energy market margins for peaking units and the outcome of the CFD structure would be to stabilize the impact of price volatility.

The CFD payment structure does, however, present design issues of its own:

§ There might be rules to specify certain hours in which the CFD offset doesn't apply (for example during outages) or it could be firm.

§ Since unit commitment is potentially a System Operator decision, the CFD offset might not apply based on ex post prices if a unit had bid a lower price but not been selected – i.e. there might be rules to limit CFD offsets over a daily cycle so as to limit some unit commitment risk.

§ The level of the strike price of the CFD must be determined. The strike price should be high enough so that the net energy margin after the CFD offset payment is positive for all generators that generate (or virtually all generators). Too low a strike price would, as discussed previously, limit efficiency in the short-term market. With all generators having the same CFD, the strike price could effectively cap the energy market prices since a unit could hedge its CFD offset exposure by bidding the strike price. A CFD payment structure would therefore contribute strongly to market power mitigation. However from the perspective of the effectiveness of short-term price signals in the energy market, it might be undesirable to have prices effectively capped at a level related to the variable cost of a peaking unit. It might be desirable therefore to adjust the CFD offset formula somewhat, perhaps by multiplying by a fixed factor set between zero and one. That level of that factor might be permanently set after an investigation of the trade-off between risk reducing/ market power mitigation benefits on one hand, and the value of efficient short-term price signals on the other.

- Finally, a decision must be made as to whether the strike price is indexed to any fuel price index. Under a CFD, a generator's energy margin is the difference between the strike price and its variable operating costs (which are largely fuel). Indexing the strike price to a fuel price index might be beneficial if it makes the margin more predictable – i.e. if it reduces generator risk.

B.3. Penalties

Penalties are closely related to payment structures. Payment structures that tie very closely to actual performance are usually not accompanied by penalty provisions. Payment structures that rely on self declaration and assumed performance in the absence of contrary evidence usually have penalty provisions.

For example, in energy markets there are generally no penalties. If a unit fails to generate, it does not get paid. However operating reserve markets often do have penalties. This is because operating reserve is very often not called to show that it can produce. A generator declares it has reserves and gets paid for reserves, usually without any short-term test of whether it could perform. In instances where reserve is called on and does not perform, a penalty applies, as it calls into doubt whether past declarations of operating reserves sold was a misrepresentation.

To the extent that the CPM payment structure makes payments based on self-declarations that are only occasionally subject to verification, penalties would be expected to apply for failure to verify the ability to perform. For example, a unit with a declared DMNC of 500MW, that tested at only 470 MW, might lose 30MW of capacity payment retroactive to the last time it demonstrated by test or actual operations the ability to produce such capacity within a period of hours.

NERA
Economic Consulting

NERA Economic Consulting
15 Stratford Place
London W1C 1BE
United Kingdom
Tel: +44 20 7659 8500
Fax: +44 20 7659 8501
www.nera.com

NERA UK Limited, registered in England and Wales, No 3974527
Registered Office: 15 Stratford Place, London W1C 1BE

MMC Marsh & McLennan Companies