CHAIRE EUROPEAN ELECTRICITY MARKETS
Fondation Paris-Dauphine

CEEM Working Paper 2014-8

FIRST PRINCIPLES, MARKET FAILURES AND ENDOGENOUS OBSOLESCENCE: THE DYNAMIC APPROACH TO CAPACITY MECHANISMS

Jan Horst KEPPLER

DAUPHINE UNIVERSITÉ PARIS

Chaire de recherche soutenue par

Rte EDF EPEXSPOT EUROPEAN POWER EXCHANGE UFE

FIRST PRINCIPLES, MARKET FAILURES AND ENDOGENOUS OBSOLESCENCE:

THE DYNAMIC APPROACH TO CAPACITY MECHANISMS[1]

Jan Horst Keppler

Chaire European Electricity Markets (CEEM), Université Paris-Dauphine

November 2014

ABSTRACT

The theoretical benchmark model arguing that competitive energy-only markets with VOLL pricing can provide sufficient levels of capacity is a coherent starting point also for discussions about capacity remuneration mechanisms (CRMs). Two types of market imperfection, both stemming from the non-storability of electricity and the resultant inelasticity of demand, however require qualification of the benchmark model and can justify CRMs. The first type of market imperfection relates to the existence of security-of-supply externalities as involuntary curbs on demand under VOLL-pricing create disutility beyond the private non-consumption of electricity. In interconnected economies, utility does not only depend on individual electricity consumption but also on the smooth consumption of others. These externalities are captured in the difference between voluntary and involuntary demand response. The second type of market imperfection relates to the asymmetric incentives for investors under imperfect information. Due to the inelasticity of demand and the lumpiness of generating equipment, investors in markets for non-storable goods will err on the side of caution, underinvesting at the margin rather than overinvesting. There exists thus not an intrinsic, general case but a time- and context-specific case for CRMs depending on the shape of the load-curve, the elasticity of demand and the availability of flexibility resources. The choice of mechanism will depend on the number of hours of potential capacity short-falls and the resulting capital-intensity of the technologies most apt to respond to them. Most importantly, well-designed CRMs will set in motion the very structural dynamics towards more elastic demand, a development that might one day make them obsolete and render the theoretical benchmark model applicable again. CRMs thus require transparent and pre-announced review mechanisms at regular intervals.

1. INTRODUCTION: THE COUNTRY-SPECIFIC AND TEMPORARY NATURE OF CAPACITY ISSUES

Discussions about the question whether deregulated energy-only electricity markets can provide adequate levels of generating capacity have not yet converged towards a generally accepted theory of optimal capacity provision in real-world electricity markets. This has forced capacity remuneration

[1] The author would like to thank Dominique Finon and Marc Bussieras for their helpful comments. Discussions at the "European Workshop on Capacity Mechanisms in EU Power Markets" in April 2013 at Université Paris-Dauphine and the 8th Session of the Research Seminar in Energy Economics in December 2013 at Paris-Sciences-Lettres on capacity mechanisms around Thomas-Olivier Léautier also provided important building blocks towards this paper. This paper has benefited from the support of the Chaire European Electricity Markets (CEEM) of the Paris-Dauphine Foundation, supported by RTE, EDF, EPEX Spot and the UFE. The views and opinions expressed in this Working Paper are those of the authors and do not necessarily reflect those of the partners of the CEEM.

mechanisms (CRMs) in the real world to advance with surprisingly little help from the theoretical literature and has created a wide divergence of views at a time when the introduction of large amounts of variable renewables lends new urgency to the issue, in particular in European electricity markets.

The principal cause for this unsatisfying state of affairs is that the theoretical benchmark model spelling out a first-best optimum for energy-only markets under VOLL-pricing is ultimately too narrow a representation of electricity markets. In other words, real-world electricity markets have a number of recurring but as of yet inadequately conceptualised features that imply policy conclusions different from those emanating from the theoretical benchmark model. This does not mean that the theoretical benchmark model is *wrong* in any logical sense but that it is incomplete.

The difficulties that the community of electricity market specialists experiences with enlarging the benchmark model are due to two principal reasons. First, capacity issues touch immediately on public goods issues. A profession traditionally steeped in the limpid logic of engineering and linear optimisation struggles to come to terms with security of supply externalities. Second, capacity issues defy any general normative approach working with *ceteris paribus* assumptions. A number of parameters such as the elasticity of demand at peak time, the structure and flexibility of the generation system, the correlation of demand with renewables production or the available interconnection capacity can all significantly impact the extent to which capacity issues can be solved by energy-only markets or require additional measures to attain politically and socially desirable levels of capacity. Questions of capacity depend on the specifics of time and space.

The present article sets out in a first step to reaffirm the validity of the theoretical benchmark model for energy-only markets under the assumptions of perfect information and absence of externalities, market power and transaction costs. In a second step, it will conceptualise two features of real-world electricity markets that challenge the conceptual benchmark model. These are (a) security of supply externalities in the presence of incomplete markets for hedging against security of supply risks and (b) asymmetric investment incentives under uncertainty in markets for non-storable commodities, where the inelasticity of demand renders the losses from overinvestment greater than the profits foregone from underinvestment. Both features limit the applicability of the theoretical benchmark model and can motivate capacity support measures.[2]

This paper is thus a contribution to the literature on the divergence between the private and the social value of capacity provision. However, rather than to postulate a general public good of "adequacy", it links this divergence to two precisely defined types of market failures which tend to manifest themselves in different electricity markets to different degrees and in different forms. It is the case-by-case characterisation of these market failures that will need to inform the precise form of the desirable capacity mechanism in each electricity system. There is thus a *tendency* towards the need for some form of capacity remuneration in competitive electricity markets. It is however

[2] The fact that the arguments for capacity mechanisms ultimately depend on externality and transaction cost arguments also explains the difficulty for energy economists to organise a more systematic and coherent debate on capacity issues. For principal methodological reasons, theoretical economics will always tend to exclude non-codifiable goods such as the security of electricity supply. However facts can be stubborn and the overwhelming empirical evidence of a looming capacity issue has forced the profession to address the issue head-on.

impossible to make the case for capacity mechanisms on the basis of first economic principles assuming full information and the absence of market failures.

The validity of the theoretical benchmark model for energy-only electricity markets is thus not an issue of principle but a question whose answer depends on the presence, degree and precise form of the two market failures mentioned. While they can be considered present today in most major electricity markets, they will manifest themselves in different forms. Successful capacity mechanisms will need to take into account the country- and region-specific nature of these market failures in their design. In matters of capacity remuneration, there is thus no one design fits all countries or situations. A mechanism for France needing large amounts of additional capacity for less than 200 hours a year will differ from Germany, which especially in the South needs to keep existing capacity designed for up to 2 000 hours per year on stand-by, or from the United Kingdom, which urgently needs large amounts of new baseload capacity. Other countries such as Norway, which has vast reserves of storable hydropower, may do very well without any additional capacity mechanism. Issues raised by different supply and demand constellations require differentiated answers in order to determine the capacity mechanisms most appropriate to address them.[3]

In addition, capacity remuneration mechanisms (CRMs) have dynamic impacts which will advance their own obsolescence by promoting the very structural changes that will reduce security of supply externalities, extend risk coverage and render underinvestment as costly as overinvestment. In short, well-designed capacity mechanisms will promote demand elasticity as well as storage, whose absence is the primary reason for a need for added capacity remuneration. Again, the precise form of such structural changes will vary from country to country.

This article thus provides a differentiated answer to the question of the need for CRMs. On the one hand, it shows that capacity mechanisms are *not* an intrinsically necessary add-on to competitive electricity markets. On the other hand, it argues that at the current stages of technology and elasticity of demand there is a strong tendency towards the provision of socially sub-optimal levels of dispatchable capacity. In other words, there are currently capacity issues in most electricity markets that need to be addressed by appropriate mechanisms. However, technological and behavioural developments already underway (think of improved load-following capabilities, enhanced interconnections, cost-effective storage, better forecasting or demand-side management etc.) suggest that the need for such capacity mechanisms may diminish over the coming decade or two.[4]

---

[3] All CRMs, whether country-specific or not, pose the question of how to organise cross-border participation and coordination at the bilateral or multilateral level. This is a thorny issue, in which the logic of mutualising of systems with different demand and supply systems competes with fears of free-riding. Treating it with the appropriate diligence is beyond d the scope of this article, which aims at providing a broader and more operational theory for capacity issues at the national than what has been available so far. The important point in this context however is that taking account of the country-specific features of the market failures that argue for CRMs has no impact on the issues pertaining to cross-border participation. For a discussion on the feasibility of cross-border cooperation between CRM see Finon (2013).

[4] The current article, which is part of a larger research project on CRMs, is not concerned with the nature of individual CRMs (long-term contracts, capacity payments, auctions, markets for physical or financial capacity options etc.) that would minimise system costs and optimise the generation of desirable behavioural and technological change. Its purpose is to provide a coherent rationale and framework for the development of CRMs according to the specific market failures prevailing in given countries in given historically determined circumstances.

Well-designed capacity mechanisms thus address the market failures that prevent equating the private and the social value of capacity and security of supply in two manners. First, in a perspective of static optimisation, they offer the additional incentives required to provide optimal amounts of capacity. Second, in a dynamic perspective, they foster the structural change that will promote the ability of energy-only markets to provide eventually desirable levels of capacity by themselves. They do so by reducing the transaction costs that have prevented the internalisation of welfare-relevant externalities in the first place. Regular procedures for review and adaptation with well-advertised timeframes and regulatory processes are thus indispensable features of any well-designed capacity mechanism.

The structure of this article is as follows. Chapter 2 of this article will briefly present the theoretical benchmark model with full information, no transaction costs and no market failures, in which competitive energy-only markets provide privately and socially optimal levels of capacity. Of course later chapters will move away from that model in order to precisely argue the case for capacity mechanisms under well-defined circumstances. However, establishing a clear benchmark allows doing away with the misconception that the under-supply of capacity in energy-only electricity markets is inevitable. Chapter 3 makes the case for capacity mechanisms on the basis of the fact that energy-only electricity markets tend to under-price capacity due to significant security-of-supply externalities and the inability of consumers to properly hedge against security of supply risks due to these external effects. Chapter 4 makes the case for CRMs on the basis of the fact that discontinuities in electricity price formation will asymmetrically induce producers to underinvest rather than to overinvest in capacity, an effect that is exacerbated by risk aversion and lumpiness of investment. Chapter 5 will conclude.

2. THE "BENCHMARK MODEL" FOR OPTIMAL CAPACITY PROVISION IN ENERGY-ONLY MARKETS

Before spelling out the rationale for dedicated capacity remuneration mechanisms (CRMs) based on specifically identifiable market failures that prevent full cost recovery at socially optimal levels of security of supply, it is useful to recap briefly the "benchmark model" (Léautier) of optimal capacity provision and full cost recovery in energy-only electricity markets (see Boiteux (1960, 1949), Stoft (2002) or Joskow (2007) for expositions). The benchmark model applies in principle both to a monopoly provider of electricity aiming at the maximisation of social welfare as well as to liberalised and competitive electricity markets with free price formation. On a level of first principles, assuming full information and no transaction costs under static optimisation, the two models are structurally identical. In practice, of course, differences pertain to dynamic incentives for efficiency gains and innovation on the one hand and to different levels of certainty for long-term industrial planning on the other.[5]

[5] In practice and in the context of the capacity issue, the optimizing monopolist presented in (Boiteux (1960, 1949)) has the advantage to internalize security of supply externalities by resorting to peak-load prices that "spread out the peaks and fill in the hollows" (p. 176), which is, of course, precisely what one would demand from a well-performing capacity mechanism. However, this argument cuts both ways. The implicit taking-into-account of all sorts of ultimately unverifiable externalities "public goods" (reaching from security of supply over social cohesion and industrial policy to satisfying particular

However, before entering into the aspects not covered by the benchmark model, its recap will be useful for three interrelated reasons. First, it shows that the benchmark model is not "wrong" in any logical sense and would work in the absence of the market failures identified in subsequent sections. Furthermore, in keeping with the general slant of this article, energy-only markets might one day work satisfactorily on their according to the standard theory once these market failures have ceased to exist due to the technological, informational or behavioural changes induced by currently required CRMs. Second, it cuts short the fashionable but misguided twaddle that considers generating capacity "a good separate from electricity", which therefore "needs its own price and market".

This is lazy thinking. Generating capacity is a fixed factor of production in electricity generation. All that CRMs provide for is a smoother time distribution of the amount of revenues that correspond to full cost recovery thus leading to less volatile electricity markets with higher levels of security of supply. In a theoretical perspective of static optimisation, economic costs with a CRM will be slightly higher than in an energy-only market with VOLL-pricing due to the subsidisation of extreme peak demand by a broader segment of the market? The latter's extent depends on the specifics of the capacity support mechanism that has been chosen. However, CRMs might well bring down total costs over time by generating better market-wide information and reducing uncertainty. Third, the benchmark model is indeed a good starting point for illustrating and exemplifying market failures existing in real-world electricity markets.

In principle, electricity is an ideal good for competitive markets. Since electricity cannot be differentiated beyond very basic, easily observable and enforceable criteria (frequency, voltage, stability), it allows the functioning of a market that outside of the world of finance constitutes the rare example of a market without transaction costs or product differentiation.[6] Unsurprisingly, strict marginal cost pricing is the norm in competitive electricity markets outside the hours of extreme peak demand. During these hours, the threat of service interruptions is supposed to increase prices to the value of lost load (VOLL), which corresponds in standard microeconomic parlance to the marginal utility of electricity.

In theory, decentralised decision-making in competitive electricity markets will provide a level of capacity such that prices at peak demand and the number of hours that they will prevail will be sufficient to allow recuperating fully all costs of production, including fixed costs. This level is, of course, considerably above the variable cost of the marginal technology.[7] During a certain number of hours, prices will thus reach a level at which the equilibrium between supply and demand is

---

political constituencies), was, of course, at its time a potent argument against the monopoly provision of electric power and in favour of electricity market liberalisation.

6 Physical network losses are precisely measurable and codifiable to the extent that they constitute a perfectly operative sub-market that allows feeding them back without any economic efficiency losses into the main market.

7 In order to avoid terminological confusion we will call "marginal technology", the power plant for generating electricity with the highest variable costs. This clarification excludes "demand response" as the marginal production technology. This is done for reasons of readability only. In principle, demand response, in particular if it is triggered by dedicated technical hardware, can be considered a marginal technology since it is this action that will be responsible for equating supply and demand at the margin. For reasons of terminological clarity, however, it seems preferable to leave "demand response", "load-shaving", "demand-side management" etc. on the other side of the supply-and-demand equation, namely demand.

established by no longer satisfying a portion of demand. In markets with inelastic demand this will be done through rolling brownouts (involuntary demand response), in markets with partially elastic demand through voluntary demand response to the extent that it is available.[8] The difference in the economic costs of involuntary and voluntary demand reduction consists precisely of the security of supply externalities that will be discussed in section 3. Typically, the number of such scarcity hours per year at which VOLL-pricing prevails is measured in the single or low double digits, while prices that at this point correspond to the marginal utility of electricity are measured in the thousands of dollars or Euros.

The precise number of VOLL hours can be determined either by the market or by the level of VOLL determined by the regulator (or the optimising monopolist) in order to avoid prices going towards infinity. In the latter case, the regulator expresses the social preferences determined through the regulatory process. In either case, investment will adapt in a manner such as to produce in the interplay with electricity demand a level of capacity that will determine a certain number of annual VOLL hours. The statistical average of these VOLL hours multiplied by the amount of VOLL, set either by the market or the regulator will correspond to the otherwise "missing money" required to recuperate the full costs of capacity. The higher the VOLL, the smaller the expected value of the number of hours during which it will reached and *vice versa*. In a full-information, no-transaction cost world with no externalities, there exists thus no need for capacity mechanisms as the sequence of balancing, intraday, day-ahead and forward markets trading in energy only will generate the appropriate incentives for socially optimal level of investments.

The following equation summarises the principle of full cost recovery under both short-term marginal cost (variable cost) pricing and long-term capacity cost (VOLL) pricing:

$$[FC_i + VC_i * h_i] * CAP_i = [\sum_m (VC_m - VC_i) * h_m) + (VOLL - VC_i) * h_{VOLL}] * CAP_i \quad \forall\, m\ with\ VC_m \geq VC_i.$$

Where,

FC_i indicates the annualised investment costs of technology i.

VC_i indicates the variable costs per unit of output of technology i.

h_i indicates the hours of operation per year of technology i.

CAP_i indicates the installed capacity of technology i.

VC_m indicates the variable costs of the marginal technology that sets the price.

h_m indicates the hours of operation per year of technology m.

$VOLL$ indicates the value of lost load, and

h_{VOLL} indicates the number of VOLL hours per year.

The condition $VC_m \geq VC_i$ indicates that technology i can itself be the marginal technology and it holds that $\sum_m h_m = h_i$. In cases, where $VC_m < VC_i$ technology i does not operate.

[8] Nobody has done a better job in recent years about educating economists about these fundamental relationships than Paul Joskow in "Competitive Electricity Markets and Investment in New Generating Capacity" (2007).

The equation above synthesises the three central features of the standard theory of optimal pricing in electricity systems, whether they are governed by the prices resulting from decentralised profit-maximisation in competitive markets or by the tariffs set by a benevolent monopolist aiming at maximising the social surplus:

1. **Short-term marginal cost pricing, which corresponds to variable cost pricing, at all times,**
2. **Full cost recovery and the satisfaction of budget constraints both at the level of the individual firm and the system in the sense that annual revenues are equal to total annual costs.**[9]
3. **Long-term marginal cost pricing during extreme peak or VOLL hours.**

The defining result of the standard theory with extreme peak pricing is that (a) due to short-term marginal cost pricing at all times it is privately and socially optimal (b) due to long-term marginal cost pricing during extreme peak (VOLL) hours all actors satisfy their budget constraint and are able to recover their full costs including fixed capital costs. This contradicts the standard economic result for firms producing under increasing returns to scale, namely that social optimality would require a combination of short-term marginal cost pricing and tax-financed subsidies in order to pay for fixed costs. So how can electricity markets deviate from this fundamental principle of Walrasian microeconomics? The answer is that in markets for non-storable services with variable demand short-run marginal cost at peak time *is* long-run marginal cost.[10] Other than electricity, one may think of markets for bandwidth or traffic pricing. Despite appearances to the contrary, the central principle of modern microeconomics that only short-run marginal cost pricing guarantees social optimality is thus preserved.

This unique result is due to the double nature of extreme peak (VOLL) prices. Prices at VOLL hours correspond both to the short-term marginal cost of not consuming electricity, which is equal to the marginal cost of making an additional unit of electricity available through demand restraint *and* the capital costs of producing an additional physical unit of electricity! As indicated in footnote 6, such demand restraint can be considered as a particular technology of electricity production with very high marginal costs and zero fixed costs. However, no additional insights are gained by such a semantic contortion. In either case, VOLL corresponds to the disutility, the marginal utility lost, of not using the marginal unit of electricity. The key economic property remains, *i.e.*, the coincidence of

---

9 Full cost recovery under VOLL, of course, also implies the absence of any "missing money", the term used to indicate that the infra-marginal rents earned in energy-only markets are insufficient to cover the fixed costs of a level of capacity that would cover demand *at all times*. Herein lies the rub. Covering demand *at all times* at prices that correspond to, say, the variable costs of the marginal technology, precisely implies the avoidance of pricing electricity at the value of lost load. Employing the term "missing money" thus implies that competitive and privately optimal levels of capacity are considered socially suboptimal. This is wrong from a purely theoretical point of view. However, as will be shown above, there may be money missing, if one supposes that for reasons of market failures capacity should be higher than that provided by the market.

10 This is different from average cost pricing in industries with increasing returns to scale producing storable commodities. Due to storability, firms in such industries must not take peak demand but total demand into account when choosing their optimal capacity. In such cases, it can be easily shown that only marginal cost pricing at variable costs of production is socially optimal. In industries with increasing returns to scale, this requires subsidies for capital costs in order to ensure economic viability.

short-term and long-term marginal costs. As expressed by Marcel Boiteux, one of the founders of the theory of peak-load pricing:

> "Under the theory of selling at marginal costs, prices must be equal to the *differential costs* [short run marginal costs] for *existing plants*. Plant is of optimum capacity when the differential cost and the development costs [long run marginal costs of additional capacity] are equal, that is to say, when differential cost pricing covers not only working expenses but also plant assessed at its development cost (Boiteux (1960, 1949), p. 167)."

How can such a unique coincidence of short-term and long-term marginal costs come about? All that is necessary is that producers are capable of adding or subtracting generating capacity to and from the market such that the product of VOLL and the number of resulting VOLL hours corresponds to the balance of their fixed costs. An interesting question arises about whether the market needs to be competitive or not in order for full cost recovery. Stoft maintains that full cost recovery in a liberalised electricity market does not depend on the market being competitive.

> "The discussion of fixed-cost recovery does not depend on any details of the cost-functions or even on the market being competitive. It depends only on the ability of generators to enter and leave the market (Stoft (2002), p. 123)."

True enough, but of course in an uncompetitive market, generators would recuperate *more* than full costs by restricting capacity and increasing VOLL hours beyond the level necessary to recuperate fixed costs. Stoft is thus not entirely correct that full-cost recovery "fails if there are barriers to entry (*ibid.*)". Full-cost recovery would still work but it would no longer arrive at socially optimal outcomes. Léautier is thus correct in making the competitiveness of electricity markets the primary condition for the absence of underinvestment, as long as other market imperfections are absent (Léautier (2013), p. 10). Needless to say, as long as Boiteux' optimising monopolist is working with an objective function aiming at the maximisation of social welfare its capacity will also be optimal in the absence of other imperfections.

In the absence of market imperfections, the benchmark model for privately and socially optimal capacity provision in energy-only markets is thus alive and well. Arguments for CRMs substituting for VOLL-pricing must thus transcend the benchmark model. We will show in the next two sections that privately optimal levels of capacity in energy-only markets can be socially suboptimal due to the under-pricing of security of supply externalities and informational asymmetries, which is the primary purpose of this article.

3. THE DEMAND-SIDE: LESS THAN SOCIALLY OPTIMAL CAPACITY PROVISION DUE TO SECURITY-OF-SUPPLY EXTERNALITIES

Any divergence from the theoretical benchmark model for energy-only markets by way of capacity remuneration mechanisms implies economic efficiency losses and must thus to be justified on the basis of market failures such externalities. Such externalities must at the very least be identified on conceptual grounds and, ideally, at best be empirically verified. The latter is easier said than done. It is in the nature of market failures or externalities, in fact it is their *raison d'être*, that their identification, measurement and costing is more difficult than for marketable goods (Keppler, 2010). Sections 3 and 4 will nevertheless develop the case for two types of market failures, which both imply the provision of socially sub-optimal levels of capacity in energy-only markets. The first case pertains to the demand-side, the second to the supply-side. One the demand-side, consumers would

prefer higher privately contracted capacity due to the existence of security-of-supply externalities. On the supply-side, private investors provide on average less than socially optimal levels of capacity even in the absence of demand-side externalities due to asymmetric investment incentives in markets for non-storable goods. This effect is exacerbated by risk aversion and the increase in volatility caused by intermittent renewables. Demand-side (section 3) and supply-side (section 4) effects do not imply each other and are additive.

Starting with the first effect, the existence of security-of-supply externalities implies that consumers and political decision-makers would like to have and would be willing to pay for higher levels of security of supply than implicitly contracted for in energy-only electricity markets with VOLL (see below). However, in real world electricity markets one does not even need to go as far as identify specific externalities. Due to their ambiguous nature, real-world VOLL-pricing implies a distinct disutility. For theoretical economists VOLL-pricing represents the moment, inevitable and necessary, when operators recoup the revenue short-fall referred to as "missing money" to cover their fixed costs. For consumers and policy-makers VOLL-pricing represents the dreaded moment when electricity prices go haywire, electricity supply is cut and faith in the working of electricity markets breaks down. In other words, even if VOLL-pricing was economically justifiable, and we will show that at the present state of technology and behaviour it is not, it may not be socially and politically sustainable. One of the key issues surrounding electricity markets is the fact that consumers, politicians and most stakeholders neither like nor accept VOLL-pricing in the hundreds or even thousands of Euros even if this is fully covered by economic theory.[11] There is thus a social disutility associated with VOLL-pricing.

It is important to understand that the dislike of VOLL is not simply an irrational whim harboured by poorly informed non-experts but that it constitutes an intuitive grasp of the challenges connected with the transposition of theoretical VOLL-pricing into practice. These challenges relate precisely to the socially suboptimal provision of capacity due to (a) security of supply externalities and (b) asymmetric incentives for investors in energy-only markers discussed in the following.

Even in the absence of security-of-supply externalities and asymmetric investment incentives, it is hardly obvious that VOLL pricing would work as indicated by theory. Scarcity pricing at VOLL is in fact a very imperfect way to provide adequate investment signals by equating prices to willingness-to-pay in the very situation when demand is reaching capacity. As Joskow (2008) points out, the extreme demand and supply tensions necessary to induce load-shedding under VOLL are frequently characterised by disequilibria or even complete market breakdown that do not lend themselves to the discovery of marginal cost of electricity provision, load shedding, willingness-to-pay or prices:

> "There are a number of wholesale market imperfections… that appear collectively to suppress spot market prices… below efficient prices during the small number of "scarcity" hours in a typical year when wholesale market prices should be very high… Since the market

[11] An integrated monopolist aiming at welfare maximisation has an intrinsic advantage here over competitive electricity markets, even when as shown in section 2 both are based on the same underlying economic principles. Paradoxically, this advantage consists in the fact that contrary to a liberalised market, the monopolist is not obliged to pursue economic welfare optimisation in a narrow sense, i.e. to practise pure VOLL pricing. It can instead integrate social preferences for less-than-VOLL prices but higher levels of security of supply in the form of less-than-VOLL but higher than marginal cost prices during a correspondingly longer number of hours. This explains why the two models are often seen as antipodes, even though they are structurally identical from an economic point of view.

also collapses in these situations, wholesale market prices are effectively zero and do not reflect consumer preferences to buy or generators' cost of supply (Joskow (2006), p. 165)."

In practice, reaching capacity limits is thus associated with market breakdown in which trades are no longer made and pricing is absent or zero. Thus even in the absence of the two market failures this article concentrates on, VOLL-pricing too often remains a virtual concept unable provide a sufficiently stable price signal allowing to equate long-term marginal cost to the marginal utility of electricity.

Security-of-supply externalities

Energy-only markets provide less than socially optimal levels of capacity due to security-of-supply externalities. As always (see Coase (1961 and 2008), Arrow (1970), Keppler (1998 and 2010), such externalities are due to transaction costs and imperfect information, which prevent the creation of a working market for the good in question. Due to the complexity of the good "security of electricity supply", which depends on social preferences, political circumstances, the state of technology, behavioural structures and a slew of other factors it is near-impossible to let energy-only markets decide on the appropriate risk of security-of-supply interruptions.[12] In a market for a non-storable good such as electricity with its obligation to organise the supply and demand balance second by second, the maintenance of security of supply is however a constant and pressing issue.

The theoretical position based on first economic principles that energy-only markets can ensure sufficient amounts of investments and adequate levels of security of electricity supply no longer holds once one identifies significant security of electricity supply externalities. Their existence can be reformulated as a situation, in which the total social cost of a supply interruption or, equivalently, the willingness of society to pay for additional security of electricity supply is higher than the cost of additional capacity.

In such a situation, the fundamental contribution of capacity mechanisms is to offer a mechanism through which social preferences for security of supply can be transformed into an explicit capacity objective, which then translates, depending on the specific capacity remuneration mechanism (CRM) chosen, into provided added remuneration to capacity providers during all hours of the year or a specified sub-set of them. In accordance with the theory of externalities, CRMs thus break down the complexity of the good "security of supply", codify it in terms of a capacity target, which translates into a socially acceptable number of hours of demand curtailment, and thus reduce market transaction costs and eliminate the market failure. CRMs are thus specific, possibly temporary, measures to codify and internalise the good "security of supply", which would otherwise be too complex for markets to handle in a socially optimal manner. In other words, CRMs overcome the transaction costs that previously impeded negotiating for optimal levels of security-of-supply. They thus perfectly exemplify Coase's fundamental insight that only in a world without transaction costs social value is necessarily maximised (Coase (1988), 158).

[12] Not only is it difficult to define properly security of supply, but it is even difficult to define a security of supply breakdown. Pierre Bornard, President of the Supervisory Board of ENTSO-E, the European association of TSOs, likes to quip that the only existing definition of a security of supply incident is the fact that the Energy Minister had to resign. Clearly, in the absence of more precise quantitative indicators for the value of security of electricity supply, the social disutility of such an event is hard to monetise.

The disconnect between social and private preferences for security of supply in the presence of transaction costs can be demonstrated by way of a simple example stemming from the French electricity market. France has always had the virtue to have an explicit security of supply target. The latter is currently set at a level of three VOLL hours per year. With prices on the French-German day-ahead electricity market EPEX Spot being institutionally capped at € 3 000 per MWh, one can easily see that VOLL pricing during three hours per year and at a French peak demand of circa 100 GW will yield at best € 900 million per year. This however is not nearly enough to recuperate "missing money" that at annualised capital costs of € 50 000 per MW for a combustion turbine would stand at € 5 billion. In reality, the situation is even more dramatic. Reaching the cap of € 3 000 per MWh for single hour in 2009 produced a political uproar and a serious questioning of the adequacy of liberalised electricity markets across the political spectrum. Traders and theoretical economists did their best to defend this as part and parcel of the working of electricity markets but were drowned out in the discussion.

Previous attempts to define security of supply as a public good were inadequate

The idea that security of electricity supply is in a yet to be defined sense a public good is not entirely new. Oren (2003), Kiessling and Gibberson (2004 and 2007) and de Vries and Hakvoort (2004) have all made this point in various forms. While all of these authors are good electricity market experts, neither makes a coherent conceptual argument why security of supply issues due to underinvestment in capacity may arise in competitive and liquid energy-only markets. This leads to circular arguments such as "CRMs are needed when energy-only markets do not work properly," which do not advance our understanding of what precisely constitutes the market failure in question what may be the role of CRMs in eliminating it. Oren's 2003 well-known paper on generation adequacy is a case in point:

> "When energy markets are not sufficiently developed to provide correct market signals for generation investment, setting capacity requirements with secondary markets that enable trading of capacity reserves is the preferred approach. It is more likely to produce correct market signals for investment than administratively set capacity payments which are likely to distort energy prices and result in over-investment (Oren (2003) p. 21)."

Not only are notions such as "not sufficiently developed" too vague but also the notion of "over investment" requires a sharper definition. From a social point of view "over investment" beyond competitively supplied levels is precisely what you want. Like a number of commentators, Oren also attempts to reduce the capacity adequacy issue to a question of indelicate behaviour by private operators:

> "An important concern that is often voiced in countries where there is no well developed institutional infrastructure that can enforce financial liability of corporation is that load serving entities or generators may assume more risk than they could handle reliably… We cannot ignore the reality that US bankruptcy laws provide a de facto hedge to load serving entities which may result in assumption of imprudent risk (Oren (2003), p. 15)."

While US bankruptcy may or may not encourage excessive risk taking, this has nothing to do with the specific coordination failure related to the socially sub-optimal provision of capacity, which at low elasticities of demand persists even with perfectly hedged and risk averse operators. Conversely,

with the right incentive structure even the most indelicate operator would provide adequate levels of capacity. There is an unfortunate tendency in discussions surrounding electricity markets to moralise structural issue. This, of course, obscures, rather than clarifies the real issues behind socially adequate capacity provision. It is Coase's great merit to have irreversibly shifted the externality issue from an unwillingness to trade (a moral issue) to an inability to trade (a structural issue).

While the reasons for socially sub-optimal investment in capacity in liberalised electricity markets are made slightly more explicit in a paper by de Vries and Hakvoort, the authors also fall back on morally doubtful "free-riding" as the primary cause for this unsatisfying state of affairs. They first provide a useful list of "factors which may disturb the narrow investment optimum. The following types of market failure can be discerned (…):

- Price restrictions,
- Imperfect information e.g., regarding consumer willingness to pay or future supply and demand,
- Regulatory uncertainty,
- Regulatory restrictions to investment, and
- Risk-averse behavior by investors (de Vries and Hakvoort (2004), p. 4)."

These points are well worth mentioning. However, de Vries and Hakvoort do not relate them properly to the distinction between private and public goods. In other words, they fail to spot the externality, although they introduce the term:

> "In a [socially optimal] market equilibrium, this positive externality would be reflected by consumers not revealing their true willingness to pay. If service interruptions are the consequence of, for instance, a 2% shortage of generation capacity, this means that service interruptions affect only about 2% of the customers at a time during a period of scarcity… The consumers who caused the shortage by under-contracting therefore do not suffer the full consequences; instead, they still can consume as much electricity as they want for 98% of the time. In a [private] market equilibrium, this means that those consumers who show a lower willingness to pay, benefit from those who show a higher willingness to pay and thereby attract more peak capacity. The public good character of reserve capacity therefore provides consumers with an incentive to understate their willingness to pay (*ibid*., p. 6-7).

This is not correct. First, security of supply externalities have nothing to do with consumers wantonly underreporting their true willingness to pay. They thus continue the argument introduced by Oren that less than socially optimal capacity is due to a minority of indelicate participants in the electricity market although they shift the issue from the supply side to the demand side. Second, at stake is not the average demand of electricity but the demand for electricity at extreme peak times, which is equal to capacity. In other words, in question is not the average willingness-to-pay for electricity but the marginal willingness-to-pay for electricity at time of scarcity. If there are consumers that under-contract their true consumption they will suffer utility losses of their own. There are no externalities nor public goods issues involved.

The public good issue was addressed head-on by Kiessling and Gibberson (2004) in their presentation on "Is Network Reliability a Public Good?" Following Oren (2003), their contribution has the merit of highlighting the fact that there are several issues involved in network reliability such as adequate

capacity provision, operational reliability and the provision of ancillary services. They also correctly point out that network reliability has both private and public good aspects.

However they subsequently set up the public good issue as a straw man to better take it down. Similarly to de Vries and Hakvoort they frame the problem in terms of heterogeneous preferences and free-riding (p. 10). The solution is then straightforward, better contracts and priority insurance (p. 19). This again misses the point. Without sufficiently elastic demand, the security of supply issue will persist even with perfectly honest, homogenous consumers as operators have no means of recuperating the full social willingness to pay for an additional unit of capacity in energy-only markets *even with fully working VOLL pricing*. In this manner no coherent argument for CRMs can be made.

What are security-of-supply externalities?

In abstract terms, CRMs become necessary if due to insufficiently elastic demand the good "security of supply" is too complex and transaction costs are too high to be traded bilaterally. Concretely, a public goods or externality issue arises if

1. The non-consumption of electricity of consumer A affects the utility of consumer B (as well as vice versa) and
2. The two are unable to move towards higher levels of capacity (A's demand for more capacity integrating B's utility and vice versa) through appropriate side-payments.

The second condition is, of course, impossible to realise without that a third codifies the good in terms of tradable capacity certificates, which means creates a capacity mechanism. Let us therefore concentrate on point 1. It is important to understand that such security-of-supply externalities arise only if the non-consumption is *involuntary*. With voluntary and possibly remunerated, demand restraint, the externalities will fade away. In other words, the underlying issue is due again to the inelasticity of demand which results from the fact electricity in most markets cannot be stored in sufficient quantities at sufficiently low cost. If the demand side was elastic, it would work exactly like storage and the public goods issue would fade away.[13]

The presence of reciprocal externalities in electricity consumption of electricity thus makes private contracting for the appropriate level of security of supply inadequate. This also means that brown outs during VOLL hours have *higher* social costs than the product of private cost (equal to VOLL) times the number of disconnected customers. The aversion of customers and politicians towards VOLL-pricing thus has a serious underlying rationale: due to network effects, the social costs of an interruption of electricity supplies are larger than the private costs. The network effects in question do not relate to the physical networks of power transmission but to the economic and social networks of modern industrial societies. Electricity pervades every aspect of society. Preventing a fraction of consumers to participate in its social and networks will inevitably propagate and thus inflict damages, real and perceived, to far larger sub-sections of the socio-economic system.

If an electricity customer, for instance, is a hospital or a restaurant, it is easy to see that the costs of even an hour's outage will affect the well-being of many other people. The question is now whether

[13] The inelasticity of the short-term electricity demand function is not only a result of technical and informational constraints but also of behavioural inertia at the level of individuals and households. All three are part and parcel of the "transaction costs" which impede the first best optimum to be realised without any externalities.

the loss of utility of a hospital's patients or a restaurant's customers – a loss of utility that can stretch over far longer periods than the actual outage – is adequately taken into account in the decisions affecting electricity supply of the manager of the hospital or the restaurant. If it is, there is no externality. If it is not, there is an externality.

A simple but incisive example may illustrate the point.[14] Imagine a visitor riding down the elevator in a multi-story office building after an afternoon meeting that stretched into the winter evening. Suddenly, the elevator stops and the lights go out due to a rolling brownout during evening peak hours. Even after electricity has come back, the stress is considerable and the evening is done for. Of course, the example can be expanded at will with a number of dramatic or hilarious ramifications. In the present context, there are two important points here:

1. **Due to the inability for the electricity distributor to single out individual customers, this situation can arise *even if the building manager has correctly anticipated both his consumption and his capacity*. This is *not* an issue of free-riding or misrepresentation of true willingness-to-pay as implied by De Vries and Hakvoort or Kiessling *et al*. This is a classic externality issue where due to high transaction costs the building manager and its tenants unable to transmit individual preferences for continuity of service.[15] This holds *a fortiori* for the hapless visitor. The good in question (security of supply) is undersupplied.**
2. **If overall electricity demand was more elastic, the building manager and its tenants might have decided to partake in a demand-side management programme which organises *voluntary* (or contracted, which amounts to the same thing) load shedding at certain peak hours. A message sent several hours before would have reminded the building manager of his obligation to minimise electricity consumption and to shut down the elevators. In this case, a warning sign "Do not use elevators" would have fully internalised the potential externality.**

The example illustrates that the security of supply externality is due to the involuntary character of the enforced load shedding. The difference between an involuntary disconnection with inelastic demand and a voluntary reduction or deferral of demand consists precisely of the positive externalities of electricity consumption. In reality, such security-of-supply externalities consist of a myriad of infinitely small impacts. Evening football matches, train and metro operations, public lighting and security as well as, ultimately, the investment climate and economic development depend on continuous, high quality electricity supply.[16]

Short of installing individual auto-generation or costly back-up systems, which are warranted only for consumers with the highest risk of massive externalities such as hospitals or data centres, individual

---

14 I would like to thank Marc Bussieras, EDF, for providing this example. He is, of course, absolved from any responsibility for the usage made of it in this context.

15 One finds here the confirmation of the principle established by Coase that the level of transaction costs determines the severity of the externality (Coase (1961)). Transaction costs impede the necessary feedback of consumer preferences to producers in form of a reliable price signal. Transaction costs increase with the informational complexity of the interactions between different actors. Given the ubiquity of electricity in modern societies, a residue of transaction costs is inevitable in a sector such as electricity. See Keppler (1998) for the link between informational complexity and externalities.

16 The straightforward models of economic theorists (see Léautier (2013), for example) treat electricity exclusively as a private good. They thus fail to make the difference between an expected voluntary and an unexpected involuntary reduction in demand.

electricity customers cannot internalise these effects into their willingness to pay for uninterrupted electricity supply. This holds for average *and* marginal willingness-to-pay, e.g. in the case of real-time metering. Even if real-time metering would include the selective disconnection of individual customers, such targeted load-shedding would not include the knock-on effects (another word for external effects) on third parties. The transaction costs to arbitrage between private willingness-to-pay and social willing ness-to-pay, e.g. in the case of security at night or "investment climate" are far too high.

In the absence of elastic demand in large swathes of the market, it will hold even with perfectly working real-time metering and tariffing arrangements in perfectly competitive energy-only market that:

- **Social willingness-to-pay for additional security of supply (additional capacity) > Private willingness-to-pay for additional security of supply (additional capacity) = private marginal cost of capacity.**

The social costs of supply disruptions thus exceed for the time being the private willingness-to-pay for energy that can be captured by producers in an energy-only market for providing capacity. Hence, the number of VOLL hours in a liberalised energy-only market will be higher than the social optimum. We say "for the time being" as transaction costs and hence the inelasticity of demand, are not fixed through time. Load shifting can function like storage, as demand rather than electrons is "stored", i.e. transferred through time. The adoption of demand technologies that make the loss of utility due to voluntary reductions of electricity consumption amenable to compensation may over time indeed reduce the gap between privately and socially optimal levels of capacity. While the gap will never be zero, it may become negligible.

4. THE SUPPLY-SIDE: LESS THAN SOCIALLY OPTIMAL CAPACITY PROVISION DUE TO ASYMMETRICAL INCENTIVES IN MARKETS FOR NON-STORABLE GOODS

Other things being equal, non-storable goods will always have more inelastic demand than storable goods. In addition to creating externality issues on the demand side, this inelasticity of demand provides also asymmetric incentives for investors in capacity on the supply side. This pushes actually provided levels of capacity further away from socially desirable levels of capacity. While the issue is also due to the inelasticity of supply (and may fade away as demand becomes more elastic), it is independent of the security of supply externalities spelled out above.

The reason is that electricity generation investments cannot be scaled to an arbitrarily fine degree. In combination with the inelasticity of demand, this means capacity investment will always either slightly over- or undershoot the theoretically optimal amount. However, the implications for profits are not symmetric for over- or underinvestment. Overinvestment creates small gains in added quantities sold and large penalties in terms of price declines, even for small amounts of excess capacity. Underinvestment creates small losses in terms of sales foregone but large gains in terms of more frequent scarcity pricing. Due to the extreme inelasticity of demand at peak time, the issue poses itself not only at the level of the industry but at the level of the individual producer.

An example illustrates the point. Assume that in a given year extreme peak demand is expected to be 101 GW for ten hours. Abstracting from security of supply externalities, one may assume that these

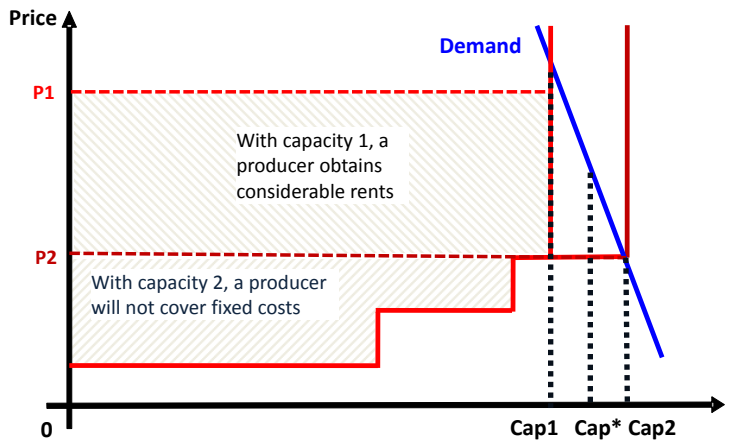
ten hours of VOLL are considered acceptable by the system operator and sufficient to recuperate the "missing money" for fixed investment costs. Assume further that the optimal system size would be 100 GW, that current capacity is 99 GW and that the minimum size of a generation investment is 2 GW. In reality the size is of course much smaller but all that is required is that it remains non-negligible with respect to the size of the market.

Any producer in the market thus has only the choice between 99 GW with 20 hours of VOLL and 101 GW with zero hours of VOLL. Demand is assumed inelastic with respect to prices and prices are equal to variable cost at € 70 per MWh if demand is below or equal to capacity and that prices are equal to VOLL at € 3 000 per MWh if potential demand is above capacity.

In a market for storable goods with elastic demand, investors in capacity would face symmetric incentives. This means they would be indifferent between their opportunity loss for underinvesting, losing out on profits not made, and overinvesting. If he underinvests, new entrants have profitable opportunities for entry. On average, the market will provide the right level of investment at 100 GW. With inelastic demand and lumpy investments, the pay-offs are for over- or underinvesting are no longer symmetric. Investors will be forced to err on the side of caution and underinvest rather than overinvest. With 99 GW they will forego profits on 1 GW but will earn VOLL during 20 hours. With 101 GW they will earn profits on 100 GW during much of the year but prices will never rise above variable cost.

This is not a problem of market concentration! Competition will not change the problem, as long as the minimum investment size remains discrete. Even in a market with perfectly free entry, a potential new competitor will not enter. In fact, he would never recuperate his missing money as with his added investment demand will always be below or equal to capacity. Market power in electricity markets is, as far as investment is concerned, a structural not a legal or a moral issue. This is due to the short-term inelasticity of demand, which in return is a function of the absence of storage, and the discrete size of economic generation capacity. The graph below illustrates the same point.

Penalties for Getting Capacity Choices Wrong are not Symmetric



The incentive to underinvest rather than to overinvest holds in an environment of perfect foresight and full information as long as capacity cannot be infinitely scaled.[17] It is exacerbated by uncertainty over final levels of demand and risk aversion. Uncertainty coupled with risk aversion works as if the minimum discrete size of investment had increased. In the example above, an investor would invest if there was certainty about demand being 100 GW outside of extreme peak hours and if it was possible to invest in 1 GW increments of capacity. Under uncertainty and risk-aversion, with an *expected* demand of 100 GW, he would no longer countenance investment of 1 GW only, because the risk of overshooting in 50% of the cases would be too costly.

The discontinuities in the pay-off function related to the inelasticity of short run demand ensure that investors will always err on the side of caution, preferring a situation of slight underinvestment to one of slight overinvestment. Of course, the logic is not infinitely extensible and once capacity falls too far, new entrants will present themselves. However, due to the discrete size of their investments, they will face the same truncated profit function as the incumbents in the sense that profits are zero for any probability that demand exceeds capacity.

This is, of course, different in industries with granular investment sizes or elastic demand. If each one of these factors was present an investor would have symmetric incentives to get as close as possible to a capacity that corresponds to expected demand, Cap* in the graph above. The electricity sector, however, combining discrete investment choices and inelastic demand, will always induce investors to lean towards underinvestment.

The insufficient contribution of short-term reserve markets

So far, we have identified two general structural issues in electricity markets, the existence of security of supply externalities and asymmetric incentives in markets with inelastic demand and discrete sizes of investment, which lead to the divergence of privately and socially optimal levels of capacity thus diverge further. These general issues are currently magnified in European electricity markets by the decrease in average prices and the increase in price volatility due to large amounts of variable renewable capacity. Current low prices lead in particular to the early decommissioning of plants whose primary function was to be available during periods of high demand and high prices, the peak- and mid-load plants with comparatively higher variable costs. Thus leads to the absurd situation, in which even informed observers can insist in the same conversation on both the existence of *overcapacity*, which is correct from the point of view of private profitability, and risks to the security of supply, which would imply *under-capacity* and which is correct from a social point of view. Realigning private and social optimality is, of course, the function of appropriate capacity mechanisms.

Even if one would consider the two general issues identified above as empirically not important enough to outweigh the transaction costs associated with a capacity mechanism, it is impossible to deny that extreme patience and tolerance for risk would currently be required to invest in dispatchable capacity in European electricity markets. In this situation, much hope for providing adequate investment incentives is placed on short-term markets for balancing and adjustment.

[17] Due to the very high inelasticity of demand at peak hours, there is practically no lower bound for the size of an industrial investment in generation capacity below which one could argue that it has no influence on the demand and supply balance.

The idea is that as market demand moves closer to capacity short-term flexibility markets will begin to approach VOLL and thus cover the short-fall in remuneration. A recent White Paper on capacity mechanisms of the European Commission states:

> It has been argued that the downward pressure on day-ahead electricity prices in some markets leaves generators exposed to insufficient returns to cover their fixed costs… However, when intraday, balancing and ancillary services markets operate efficiently, such plants [mid-range and peaking] can participate in those markets, deriving additional revenue… Prices in those markets should be allowed to raise [sic] above short run marginal cost, enabling generators to cover also part of their fixed costs (European Commission (2013), p. 13).

The trouble is not only that average prices in these markets are not significantly higher than in the day-ahead and future markets but also that their price volatility is higher and hence the implied equivalent value for risk-averse operators is lower.[18] The two tables below provide an indication of the orders of magnitude involved in the French balancing and adjustment market.

Risk Aversion and Required Levels of Compensation (Balancing Market)
(RTE "marché d'ajustement", first eight months of 2013, EUR per MWh,
standard deviation = 27.80 €/MWh)

Level of risk aversion	**Average revenue**	**Effective revenue considering risk aversion**
Risk neutrality	32.47	32.47
Constant relative risk aversion (CRRA) = 1	32.47	13.72
Constant relative risk aversion (CRRA) = 2	32.47	51.22

[18] Measures of risk aversion take their origin from the work by Arrow and Pratt who define constant relative risk aversion (CRRA) as $\beta = \mu * (-U''/U')$, where β is the coefficient of CRRA, μ is the average pay-out and U is a continuous, twice differentiable utility function. The utility function is then defined as $U = (\mu^{1-\beta})/(1-\beta)$, which collapses to $U = \ln \mu$ for $\beta = 1$. CRRA has the advantage over constant absolute risk aversion (CARA) that it takes into account wealth effects, *i.e.* the risk aversion for constant sums at risk declines with increasing income or μ. A risk aversion coefficient of 1 is considered a lower bound for the average investor.

Risk Aversion and Required Levels of Compensation (Adjustment Market)
(RTE "marché d'écarts", first eight months of 2013, EUR per MWh,
standard deviation = 30.51 €/MWh)

Level of risk aversion	Average revenue	Effective revenue considering risk aversion
Risk neutrality	40.84	40.84
Constant relative risk aversion (CRRA) = 1	40.84	20.26
Constant relative risk aversion (CRRA) = 2	40.84	61.42

In comparison, during the first eight months of 2013, the EPEX Spot electricity market displayed an average day-ahead price (μ_{2013}) of EUR 42.18 per MWh with a standard deviation (σ_{2013}) during this period 17.21. In order to provide an effective contribution to the financing of new capacity investments, prices would need to be considerably higher and volatility below. At current levels, which are even below the marginal costs of conventional thermal power plants, balancing and adjustment markets may provide appropriate short-term signals for dispatch but make strictly no contribution to capacity finance.

Critics of CRMs will argue that the statistical evidence reported above is an indicator that France currently does not experience any capacity constraints. They should think again. In the two first weeks of 2013, Europe in general and France in particular, came very close to a serious supply shortage due to a strong cold spell. Experts believe that with the additional closure of roughly 10 GW of gas-fired capacity in continental Europe in the meantime, system operators would be hard pressed to maintain the demand and supply balance in a similar situation. In other words, the electricity system was at its very limit during this period.

However during the period from 3 to 16 February, prices in the EPEX Spot day-ahead market only rose to a volume-weighted average of 110 Euros per MWh, despite reaching impressive hourly peaks of 1939 and 605 EUR/MWh on 9 and 10 February. During the period of 3 to 16 February, a total of 2.7 TWh was traded. This is slightly above the amount to be expected for a two-week period, but only around 0.5% of France's total annual electricity consumption. With an average price of EUR 47 per MWh, electricity producers thus gained a surplus of EUR 170 million Euros (the difference between prices during the cold spell and average prices times the traded quantity) during the most severe supply crunch in recent history. These EUR 170 million were a welcome top-up for France's electricity producers but with EUR 1 700 per MW of installed capacity, the amount is at least two orders of magnitude (!) too low to make the slightest difference to France's installed capacity of around 100 GW. In summary, the most serious supply and capacity crunch in recent history was a very long way off of providing the sort of financial incentives for capacity development that proponents of VOLL pricing take as a given.

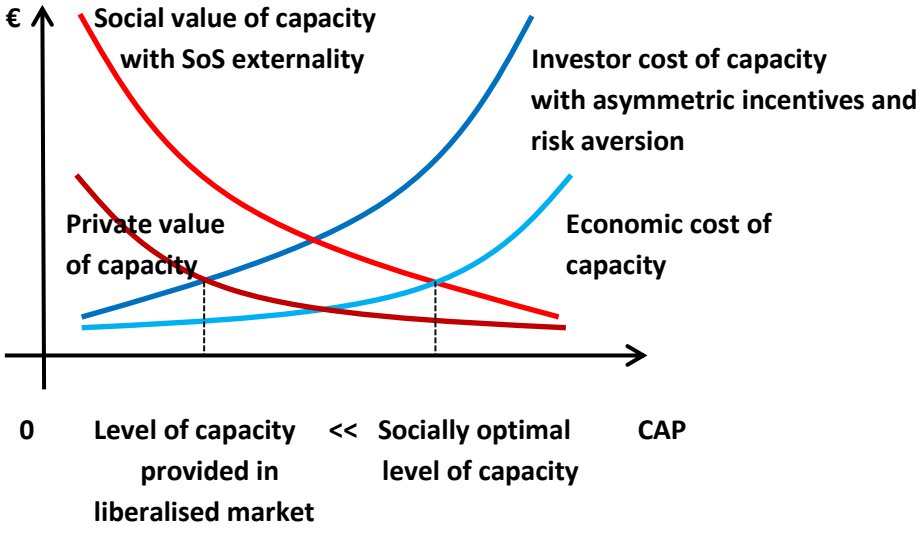
The implications of this are quite startling: even if VOLL-pricing was working properly and was politically and socially acceptable (in other words, if security-of-supply externalities were absent), the

capacity choices emanating from liberalised electricity markets, would still be below socially desirable levels due the peculiarities of price formation and the high levels of risk and uncertainty in electricity markets.

5. CONCLUDING REMARKS ON CAPACITY MECHANISMS

Combining the effect of asymmetric incentives magnified by risk aversion and price volatility presented in Chapter 4 with effect of the security-of-supply externalities presented in Chapter 3 provides a synthesis of the demand and supply situation for capacity in liberalised electricity markets in the graph below. Due to security-of-supply externalities societies will demand higher levels of capacity than even perfect markets with risk-neutral investors would provide. Due to asymmetric incentives and risk aversion investors would provide even less capacity than societies without security-of-supply externalities would demand. These two forces are *not* the two sides of the same coin. They create independent and additive effects that will cause socially desirable levels of capacity to fall short of privately supplied levels of capacity in liberalised electricity markets.

Optimal and Actual Levels of Capacity in Liberalised Electricity Markets



As soon as significant capacity shortfalls are identified in energy-only markets, CRMs become the appropriate tool to ensure the provision of socially optimal levels of investment. While a range of different capacity mechanisms can be envisioned in practice, the principle behind the effort to achieve the socially rather than the privately optimal level of capacity is the same as for any other public goods issue:

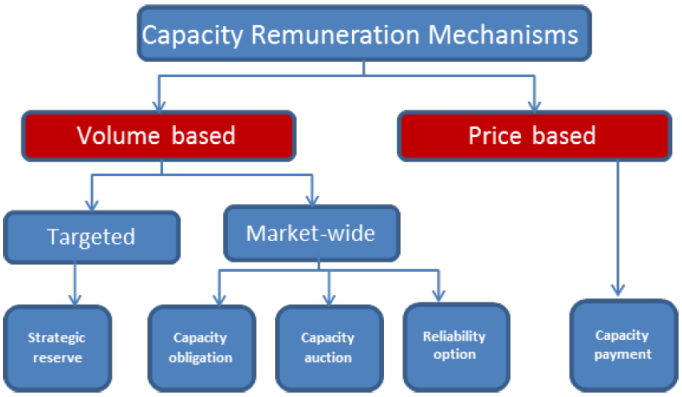
$$\sum_{i=0}^{n} MRS = MRT.$$

The sum of the marginal rates of substitution (the value of the public good of additional security of electricity supply) must equate the marginal rate of transformation that is the incremental cost of the additional capacity required. In short, producers must receive additional funds for capacity provision (Samuelson (1954)). If the security of electricity supply was indeed considered a pure public good

essential for economic development and social well-being, one could well imagine financing additional capacity payments through general taxes. In practice, political expediency will require that any additional remuneration for producers should be sourced from electricity consumers.

There exists a broad range of possible CRMs (see figure below). The analytically simplest form of capacity mechanism is a surcharge per MWh (a security of supply levy, so to say) from all customers whose receipts would then be redistributed to all generators (incumbents or newcomers) capable of offering verifiable capacity at all times. Technologically and meteorologically determined statistical adjustments need to determine the effective "capacity credit" of every installation. Conceptually, a capacity contribution would thus be analogue to a network tariff financing a shared physical infrastructure.

A Taxonomy of Capacity Remuneration Mechanisms (CRMs)



Source: ACER (2013), p. 5.

Such a capacity surcharge, whether implicit in higher electricity prices in the case of capacity obligations or explicit in capacity payments levied by the network operator will inevitably constitute an additional (usually modest) cost to consumers. In addition, from the point of view of pure theory structured around a functioning contingent of VOLL-hours which includes well codified and marketable commodities in its discourse, such extra capacity payments will necessarily count as "inefficient". In fact, they imply the "cross-subsidisation" of peak load consumption – consumption that would have been cut off during VOLL hours – through baseload consumption.

However, as long as the security-of-supply externalities have been correctly identified, the advantages in terms of reduced VOLL hours and increased security of electricity supply even in harsh meteorological conditions or unusual demand constellations will more than outweigh the costs at the level of each individual consumer or the system as a whole. In principle, the simple framework of a per MWh capacity surcharge to finance additional capacity would thus suffice to internalise the security of supply externalities connected with the VOLL hours generated by an electricity system relying exclusively on an energy-only market. In practice, three complicating factors must be considered:

Non-linear capacity surcharges: Not all customers are responsible for capacity shortages and VOLL hours in the same measure. In principle, only consumption at extreme peak periods leads to capacity shortages and should thus contribute over-proportionally to the financing of capacity. This is the very essence of VOLL pricing, which once more has theory on its side. However, between VOLL pricing,

which may equate with market breakdown and a uniform capacity surcharge, one may consider differentiated surcharges with more moderate peak pricing. Winter evening hours in France or summer middays in the United States deserve to contribute more heavily than night-time consumption in both. Each deviation from VOLL pricing will remain "inefficient" from a theoretical point of view that excludes externalities. However, a pragmatic middle way, between uniform capacity surcharges and VOLL pricing remains advisable.

The choice of the appropriate capacity mechanism will depend on the number of hours with potential capacity shortfalls

Capacity mechanisms may consist of direct payments to producers on the basis of capacity surcharges or of forward markets for tradable capacity obligations. The advantages and drawbacks are comparable to those of a carbon tax (or rather a subsidy for carbon abatement) and a carbon market. It is important to understand that "capacity markets" are markets created by regulatory fiat through an exogenously imposed constraint. "Capacity" in a capacity market – which, as has been shown above, from a theoretical point of view is excess capacity – is not a spontaneously arising autonomous commodity like electricity but a contribution to the public good of security of electricity supply. Not unlike carbon markets, capacity markets will have to struggle with concerns about credibility, long-term commitment and regulatory risk. As always, markets are an excellent means for cost discovery, however their unpredictability might not always provide the stability investors and consumers are looking for.

The answer will vary widely from country to country and hence different countries will require different capacity mechanisms. A capacity mechanism in France must find an answer to the extreme dependence of French peak demand on the weather due to the importance of electrical heating. This means that a capacity issue exists for specific hours at very cold temperatures in winter for a limited number of hours, say, even in a very cold year not more than 200 hours. In other words, the French situation calls for technologies with low-fixed costs and high or even very high variable costs, which makes, for instance, "peak shaving" through industrial demand-side management solicited through capacity obligations a promising option.

A capacity mechanism in German instead must strive towards creating conditions for secure electricity supplies all year round in the face of large-scale intermittency to the large amounts of electricity produced by more than 70 GW of variable renewables (wind and solar PV). This causes potential capacity shortfalls in the high hundreds or lower thousands of hours per year especially in Southern Germany. In this situation, capacity payments financed through auctions sufficient for allowing CCGTs to stay in business are the way to go.[19] Countries such as the UK with significant

19 There exists the alternative proposal to split the German electricity market into two different bidding zones with two different prices. The argument goes that higher prices in Southern Germany would then attract the required investment. Even abstracting from the principal issue that this would run counter to the European objective of market integration, the argument assumes that the expectation of gaining on average a few Euros per MWh more would be sufficient to maintain system-relevant producers in the system. Larger price differentials are unlikely when considering the price differentials between the German market and prices in adjacent countries (see Keppler, Phan, Le pen and Boureau, 2014). Interconnections between North and South German are despite their congestions still far larger than the interconnections with neighbouring countries. However without very substantial prices increases, the gap between prices and investment costs remains far too large and an institutional ad

needs for new baseload capacity are right to continue with feed-in-tariffs (FITs) or contracts-for-difference (CfDs) which in economic terms are very similar. No other mechanism could provide investors with the confidence that the massive funds required for 10 GW of baseload capacity will actually be forthcoming after construction periods of five years or more.

The number of hours per year over which a capacity shortfall can be expected and the resulting technologies will thus drive the appropriate choice of mechanisms without that the regulator will have to "focus" his tender on specific technologies in form of an ex ante selection. Few hours per year will mean technologies with low capital costs and market mechanisms based on obligations. A middling number of hours per year will mean auction mechanisms and a high number of hours per year will mean price guarantees for each unit of output.

There are questions here about terminology without that this must detract from the underlying purpose. A system demanding adjustments during only 200 hours per year will generate a market for "flexibility provision" closely integrated with balancing markets. CfDs are not normally thought of as capacity remuneration mechanisms, since capacity and energy are so closely related at high load factors, and yet that is what they are.

Capacity mechanisms advance their own obsolescence as they are part of a dynamic process of change towards more flexible demand and supply

Arguments for capacity remuneration mechanism along the lines developed *supra* rest on the existence of market imperfections such as externalities or imperfect information. Since Coase (1960) we know that these problems can be reformulated in terms of transaction costs. In the case of security of supply externalities, transactions costs prevent the development of products which not only insure electricity customers against supply interruptions but more importantly of products which insure everybody against the potential supply interruptions hitting everybody else. The economic rather than electric network requires the continuous functioning of the electricity system. Call it business climate, state of the infrastructure, but repeated periods of VOLL pricing, which are akin to supply interruptions for part of the population are incompatible with a competitive economy.

As for the uncertainty surrounding the precise impact of an additional unit of capacity on electricity prices that leads to asymmetric incentives for investors it is due to imperfect information another form of transaction costs. It is intrinsic to markets for non-storable goods or with high inelasticity of demand.[20]

Coase also provides, implicitly rather than explicitly, also new rationales for government intervention. The role of government is no longer to supplant the market, as is the case in all Pigouvian approaches addressing market failures but rather to *lower transaction costs* so that market itself can resolve its shortcomings (see Keppler (2010) for more details). This is precisely what happens in the case of capacity remuneration mechanisms. CRMs provide a framework, an organised

---

hoc measure such as market splitting would not make a sizeable difference given the flood of low-cost renewable energy coming being produced in both Northern and Southern Germany.

[20] Non-storability and inelasticity of demand, of course, imply each other mutually. If electricity could be stored at the level of the consumer, market demand would be much more elastic. Vice versa, demand response, which means modulating or deferring demand through time – the very essence of storage, has precisely the same impact as storage at the level of the system.

marketplace in fact, for different capacity products including demand response, short-term flexibility provision or medium- and long-term generating capacity. Theoretically most interesting are demand response and short-term flexibility provision. CRMs incentivise technologies and behaviour that render the demand curve more elastic! In this manner, capacity mechanisms are the principal tool for making the theoretical model based on forced outage during VOLL hours a reality rather than an abstract mirage that remains inacceptable to consumers and policy-makers.

In particular industrial demand-side management with clearly identifiable costs will in the future provide the sort of statistically treatable price signal that transforms uncertainty into risk and will allow socially optimal levels of investment to arise. Capacity remuneration mechanisms will thus introduce learning effects and behavioural adjustments that will hasten the very moment when peak-load pricing with *voluntary* demand response will become an economic reality. We recall that the externality element of supply interruptions resides precisely in the difference between voluntary and involuntary demand response.

The value of a CRM is thus always double: assure the correct level of security of supply in the here and now and provide the appropriate incentive for the structural change towards more demand elasticity in the future. This precisely is the reason why CRM may well be of a temporary nature as they provide incentives for their own obsolescence. This, by the way, is one of the obstacles to an easy theorisation. More important in our context is the fact that much of this induced technological and behavioural change is likely to be irreversible. An industrial consumer who invests into the technical infrastructure and behavioural skills for demand side management (DSM) in the context of a centralised capacity market will not lose these skills once the centralised market closes. In fact, he will be able to offer the same service in an ordinary adjustment market, which previously had not provided the stability and certainty for him to undertake the initial investment. There is a ratchet effect here.

This, however, has important implications for the temporary nature of capacity mechanisms. By promoting the very technological and behavioural changes that make electricity demand curves more elastic, capacity mechanisms have a tendency to render themselves obsolete. Of course, this regards long-term effects in the order of magnitude of a decade or more. Nevertheless, CRMs are addressing market failures that are not an intrinsic fatality connected with electricity markets. Of course, electricity will remain difficult to store for the foreseeable future and non-storability is the fundamental cause of the capacity issue. However, storage itself, demand response, flexible back-up, better interconnections all constitute tentative ways around the storability issues which are incentivised by capacity mechanisms.

This has two major implications. First, on a theoretical level, the question cannot be “CRMs, yes or no?” The question of introducing a capacity mechanism must depend on a serene assessment of the shape of the load curve, elasticity of the demand curve and flexibility resources. Second, as the elasticity of the demand curve and the availability of flexibility resources will be affected by the CRM in question, its way of functioning and the very rationale for its existence must be regularly assessed. Not only will one size not fit all countries, France is not Norway. The United Kingdom is not Germany or Spain. More importantly in the context of our discussion, one size will not fit the same country at different points in time. Regular, transparent and pre-announced reviews are thus an indispensable feature of any well-conceived capacity mechanism. Precisely because CRMs are dynamic in nature, as well as time- and context specific, they should be as simple and as robust as possible in order to

allow in a meaningful way for regular revisions, whose rhythm, process and decision-making criteria are well spelled out in advance.

Theoretically inclined energy economists arguing on the basis of first principles for energy-only markets and practical considerations arguing for capacity mechanisms are not in contradiction. Any good theory will survive the shock with the real world as long the underlying hypotheses, most notably, the absence of market failures, are correctly spelled out and the theory is adapted where these hypotheses no longer apply. For the time being, many electricity markets still exhibit important market failures linked to insufficiently elastic demand curves. These market failures induce private decision-makers to provide less than socially optimal levels of capacity. Unannounced and undesired disconnection during VOLL hours does not have the same economic effects as voluntary demand response. Capacity remuneration mechanism will address both the impact (in a logic of static optimisation) and the source (in a logic of inducing structural change) of the market failure and are thus an example for a policy instrument which in the long-run will create the condition, in which they might no longer be indispensable.

Bibliography

ACER (2013), *Capacity Remuneration Mechanisms and the Internal Market for Electricity*, Report 30 July 2013, available at http://www.acer.europa.eu.

Arrow, Kenneth J., (1969, 1977), "The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Nonmarket Allocation" in Robert Haveman and Julius Margolis (eds.), *Public Expenditure and Policy Analysis*. Boston: Houghton Mifflin, p. 67-81.

Boiteux, Marcel (1960, 1949), "Peak-Load Pricing", *The Journal of Business* 33(2), p. 157-179, translation of "La tarification des demandes en pointe: application de la théorie de la vente au coût marginal", *Revue générale de l'électricité* 58, *p.* 321-40.

Coase, Ronald H (1988), *The Firm, the Market and the Law*, Chicago: University of Chicago Press, 1988.

De Vries, L. J. and R. A. Hakvoort (2004), "The Question of Generation Adequacy in Liberalized Electricity Markets", FEEM Nota di lavoro 2004.120, also at http://www.feem.it/ .

Finon, Dominique (2013), "Can We Reconcile Different Capacity Adequacy Policies with an Integrated Electricity Market?", CEEM Working Paper 2013-5, Paris also at http://www.ceem-dauphine.org.

Finon, Dominique and Valérie Pignon (2008), "Electricity and Long-Term Capacity Adequacy: The Quest for Regulatory Mechanism Compatible with Electricity Market", *Utilities Policy*, 16(3): 2–14.

Finon, Dominique and Fabien Roques (2013), "European Electricity Market Reforms: The 'Visible Hand' of Public Coordination", *Journal of Economics of Energy & Environmental Policy* 2(2): 106-124.

Joskow, Paul L. (2006), "Capacity payments in imperfect electricity markets: Need and design", *Utilities Policy* (16)3: 159-170.

Joskow, Paul L. (2007), "Competitive Electricity Markets and Investment in New Generating Capacity", in Dieter Helm (ed.), *The New Energy Paradigm*, Oxford University Press, pp. 76-121 also at http://economics.mit.edu/files/1190.

Keppler, Jan Horst (1998), "Externalities, Fixed Costs and Information", *Kyklos* 52 (4): 547-563.

Keppler, Jan Horst (2010), "Going with Coase against Coase: The Dynamic Approach to the Internalization of External Effects", in J. M. Lasry and D. Fessler, *The Economics and Finance of Sustainable Development*, Economica, Paris, p. 118-138.

Keppler, Jan Horst, Sebastian Phan, Yannick Le Pen and Charlotte Boureau (2014), "The Impact of Intermittent Renewable Production and Market Coupling on the Convergence of French and German Electricity Prices", CEEM Working Paper 2014/7.

Kiessling, Lynne and Michael Giberson (2004), "Is Network Reliability a Public Good?", Presentation at 24th Annual North American Conference of the USAEE/IAEE, 8-10 July 2004, Washington DC.

Salies, Evens, Lynne Kiessling and Michael Giberson (2007), "L'électricité est-elle un bien public? *Revue de l'OFCE* 101, p. 399-420.

Knight, Frank (1921), *Risk, Uncertainty and Profit*, Boston: Houghton Mifflin

Léautier, Thomas-Olivier (2013), "The Visible Hand: Ensuring Optimal Investment in Electric Power Generation", IDEI Working Paper 605, http://idei.fr/display.php?a=22628 .

Oren, Shmuel (2003), "Ensuring generation adequacy in competitive electricity markets", University of California Energy Institute, Working Paper *EPE-007*, June 2003.

Roques, Fabien, Federico Ferrario and Philippe Vassilopoulos (2012), "Gas-Fired Power Plants on Life Support", Paris, IHS Cera, http://www.ihs.com/products/cera/energy-report.aspx?id=1065973362.

Rothwell, Geoffrey (2006), “A Real Options Approach to Evaluating New Nuclear Power Plants”, *The Energy Journal* 27(1), p. 37-53.

Samuelson, Paul A. (1954), “The Pure Theory of Public Expenditure”, *The Review of Economics and Statistics*, 36(4), pp. 387-389.

Stoft, Steven (2002), *Power System Economics*, Piscataway (NJ), IEEE Press.