Economics and design of capacity markets for the power sector

Peter Cramton and Axel Ockenfels*

30 May 2011

Contents

* Peter Cramton: Department of Economics, University of Maryland, USA. Axel Ockenfels: Department of Economics, University of Cologne, Germany. We are grateful to Steven Stoft for his important contributions to this work. We thank Matthias Janssen, Wynne Jones, and Christoph Riechmann of Frontier Economics for helpful comments. We thank RWE for funding this research and RWE staff for helpful comments. The views expressed here are our own.

1. Summary

Capacity markets are a means to assure resource adequacy. The need for a capacity market stems from several market failures the most prominent of which is the absence of a robust demand-side. Limited demand response makes market clearing problematic in times of scarcity. We present the economic motivation for a capacity market, present one specific market design that utilizes the best design features from various resource adequacy approaches analyzed in the literature, and we discuss other instruments to deal with the problems. We then discuss the suitability of the market for Europe and Germany in particular.

The capacity market we propose is an economically-sound and tested (Northeast United States and Colombia) response to the resource adequacy problem, as well as various related challenges of modern electricity markets. It takes into account the analyses of numerous proposals made in the academic and applied literature as well as many lessons learned from failed and successful attempts to deal with the problem in practice. In this way, our proposal constitutes a useful benchmark against which other proposals can be evaluated. In particular, we strongly believe that any simpler model to deal with the resource adequacy problem is likely to be economically incoherent and to hamper long-run efficiency, and we will briefly hint at examples in this paper. However, our analysis also shows for the German context, there are several other issues related to resource adequacy that need to be addressed *before* a capacity market should be adopted.

The basic purpose of a capacity market is to buy enough generating capacity in advance to assure that there will be sufficient resources when they are most needed. The ideal market coordinates new entry through the forward procurement of reliability options—physical capacity bundled with a financial option to supply energy at spot prices above a strike price. The market assures adequate generating resources and prices capacity from the bids of competitive new entry in an auction. Here, "new entry" should be interpreted in a broad sense that includes the substantial retrofitting of existing resources to address, say, environmental concerns. Indeed, the marginal capacity price may well be determined from the going forward cost of aging capacity that is required to stay online temporarily before suitable renewable or other capacity or transmission can be brought online. Efficient performance incentives are maintained from a load-following obligation to supply energy above the strike price. The capacity payment hedges load from high spot prices and reduces supplier risk as well by replacing peak energy rents (the rents derived from selling energy at very high spot prices during periods of scarcity) with a constant capacity payment. At the same time, spot prices can be as volatile as is required for short-run economic efficiency, as all parties (including load) are exposed to the spot price on the margin. Market power that would emerge in times of scarcity in the spot market is reduced, since suppliers enter the spot market with a nearly balanced position whenever the spot price is above the strike price. By the same token, this can keep to a minimum regulatory market interventions that limit the spot price during scarcity—which can be another cause for inadequate resources. Market power in the reliability market is addressed by not allowing existing supply to impact the capacity price.

While we can sketch a sensible capacity market design, our economic analysis also shows that *there are many issues that seem important in Germany that urge caution in pursuing a capacity market*. For one, while the virtues of a well-designed capacity market are readily seen in growing and stable electricity markets with a strong spot market foundation, even in these markets the development of a sound capacity market has proven difficult. One reason is regulatory imperfections: there is a history of flawed capacity market designs, resulting in large inefficiencies and costs. These flaws are well-understood theoretically, so policy makers and administrators can avoid them, but often policy makers and administrators appear guided by political forces that are vulnerable to the adoption of flawed approaches.

Second, our analysis makes clear not only the benefits of a capacity market but also what a capacity market *cannot* do. In particular, resource adequacy is a long-term challenge. That is, the proposed capacity market can unfold its benefits only if it will be in place for several decades. On the other hand, a major motivation for considering a capacity market in Germany seems to be the current transition away from nuclear power and towards renewables. (On 30 May 2011, Germany committed to shutting down eight nuclear plants immediately and the remaining nine by 2022.) That is, Germany might not look for long-term solutions of the resource adequacy problem, but rather for targeted measures to supplement the current transition phase. Such measures might include singular short- or medium-term instruments (such as "command and control") in the electricity market to avoid imminent blackouts and other emergencies due to erratic growth of renewables and insufficient transmission. They may come at significant costs, however, because such interventions obscure the actual source of the capacity problem such as lack of locational pricing, inadequate power reserve products, etc. (and provide no sensitive answer to the long-term resource adequacy problem).

Third, a capacity market does little to protect against the harm caused by politically induced uncertainties. The German market appears especially vulnerable at this time to political uncertainties regarding environmental objectives. Until these uncertainties are resolved, a capacity market is apt to be a costly and ineffective solution to the challenges of electricity investment.

Fourth, a sound capacity market requires the foundation of a sound wholesale market framework for both spot and medium-term markets. We are especially concerned by the apparent emergence of locational pricing concerns in the spot market—too little generation in the south. Locational concerns in the spot market must be resolved either with locational marginal pricing or ample transmission for a capacity market to prove effective. Otherwise, generation investments will inevitably be built in the wrong locations.

Fifth, the German market faces challenges with market integration of renewable resources that are currently out-of-market and supported by a collection of subsidies that are largely inconsistent with an efficient capacity market.

It is desirable to firmly address these issues *before* a capacity market is adopted. No capacity market can function well if there are impediments to long-term investment, such as political uncertainties, regulatory imperfections causing poor implementation, insufficient development of locational and real-time pricing, etc.

Our recommendation is therefore to first encourage long-term investment by creating a stable and sound market framework that reduces uncertainty to the extent possible, especially political uncertainty. There are also other, less invasive market design enhancements that may effectively improve scarcity pricing and resource adequacy. These concern issues like market integration of renewables, real-time flexibility of both demand and supply, further development of reserve power markets, and transmission expansion. We particularly recommend to address the chief market failure—the absence of a robust demand side—that is the primary motivation for capacity markets.

The need for a capacity market is partly reduced as the current demand-side flaws are addressed by new technology and improved electricity market designs. And the future of demand response is becoming brighter with each year as smart metering and smart grid solutions are rapidly developing. This development is likely to accelerate as consumers shift to electric cars, which will introduce a vast array of electricity storage when cars are plugged into the grid.

Embedded in a sound market and reliable political framework, a well-designed capacity market as sketched in this paper can then, *in a second step*, unfold its important complementary value in assuring resource adequacy.

2. Economic motivation for capacity markets[1]

Capacity markets are one method of assuring resource adequacy in restructured electricity markets. In this section and in section 3 we focus exclusively on capacity markets. However, we broaden the discussion in section 4. The broader discussion may be especially relevant in Germany given the rapid expansion of renewable resources and the pressures that greater reliance on wind and solar bring to the electricity system.

The liberalization of the electricity markets in Europe has mostly evolved in a period with excess generation capacity. Thus, the question whether there are appropriate incentives to invest in new generation capacity was not an issue for quite some time in most parts of Europe. This is changing as demand increases, older plants retire and investments become increasingly risky due to increased price volatility and a lack of a stable market framework. Many players and policymakers are now expressing a concern that liberalized electricity markets might not guarantee resource adequacy, and thus might fail to reliably provide electricity in the future. In Germany, such concerns are particularly fueled by an ongoing strong growth of the share of renewables. Because neither wind nor the sun can provide firm energy, renewables can only partly substitute conventional resources (such as coal and gas plants). At the same time, however, renewables increase price volatility, tend to reduce market price levels and worsen the capacity utilization of conventional capacity. This makes investments in conventional resources less attractive. Also, politics, regulation and society are constantly arguing over the right time, the right place, the right technologies and the right price for new capacity. The resulting political and regulatory uncertainties, too, may deter new investments.

Reliability of electricity supply is a particularly tricky issue in electricity markets in part because the short-term is very short and the long-term is very long: A shortage of supply on any given day can trigger costly blackouts, but it takes years to build new conventional capacity. Moreover, there are large uncertainties both on the supply and demand side that make balancing supply and demand at every instant and at every location a challenge. To deal with this challenge, most power systems established *short-term* spot (hourly) markets, complemented by power reserve markets that balance supply and demand at every second. In addition, there are *medium-term* forward (1-4 years) markets

[1] This and the other sections build on—and are partly taken from—earlier work, including Cramton and Stoft (2005, 2006, 2008, 2010), Joskow (2007), Stoft (2002), Ockenfels (2007a, 2008b), as well as on practical experiences in designing capacity, reserve and other energy markets. A German summary of this work can be found in Cramton and Ockenfels (2011).

that are supposed to provide security and firmness of electricity supply. These markets are either organized centrally or develop organically from market forces. But if there is not enough generating capacity in the *long-term*, it will not be possible to serve all load and achieve security and firmness in the short-term. In this way, adequate generation is the most fundamental reliability issue.

This paper focuses on this long-term issue of generation adequacy. It considers a centrally-organized long-term capacity market (4-20 years) that complements the short- and medium-term markets to provide adequate quantity and the right mix of resources. Part of the reason is that, while European electricity markets have not yet seen a full investment cycle, experiences from other markets indicate that liberalized electricity markets are sometimes plagued by a "missing money" problem, boom-bust investment cycles and various risks that may cause inefficient investment levels and too little reliability. A capacity market, if designed properly, may facilitate the solution to the resource adequacy problem as well as to some other threats of contemporary electricity markets, including inefficient pricing and market power. Other instruments to deal with these problems are also discussed.

Some observers argue, however, that there is no need to worry; the market and the price mechanism will take care of reliability. In fact, most other sectors do not have capacity markets. Volkswagen, for instance, does not get paid for building car factories, but for selling cars. In order to be profitable, the revenue from selling the cars must cover total costs, including the investment costs for building the production capacities. Textbook economics suggests that a competitive car market that only pays for cars works fine without capacity markets. In fact, capacity markets are typically not discussed in textbook economics at all. So, what is the economic rationale behind the idea of capacity markets in the power sector?

An economically sound motivation for a capacity market must argue that a capacity market can effectively mitigate market failures that would occur without it. Section 2 provides such arguments. The arguments are based on market imperfections and institutional constraints that are specific to the electricity sector, and not present in the car and most other industries. The described imperfections imply insufficient or inefficient incentives to build capacity and to provide reliability in electricity markets ("missing money"), possibly leading to underinvestment in generating capacity. Section 3 sketches how one can design a solution to these problems, taking into account various complexities. Section 4 discusses market-exogenous factors that may mitigate the effectiveness of a capacity market, such as excess capacity in a transition phase to a renewable-dominated system, and opportunistic political and regulatory behavior. Section 5 concludes.

2.1. Absence of demand response: what markets cannot do

An electricity market with sufficient demand elasticity always clears. That is, there always exists a market price such that demand and supply are balanced. If supply gets scarce, the price goes up until there is sufficient voluntary load reduction to absorb the scarcity. This means that, with sufficient demand elasticity, there is no capacity adequacy problem (if the market functions well otherwise): Involuntary load shedding can only occur when demand exceeds supply at all prices and so the market fails to clear. Consequently, in a market that always clears, energy prices do not and cannot determine the level of reliability for, with respect to adequacy, the market is *always fully reliable*. Instead, markets determine the *efficient* level of capacity. Any less capacity would cause expensive voluntary load reductions, and any more capacity would mean that too much had been spent on capacity.

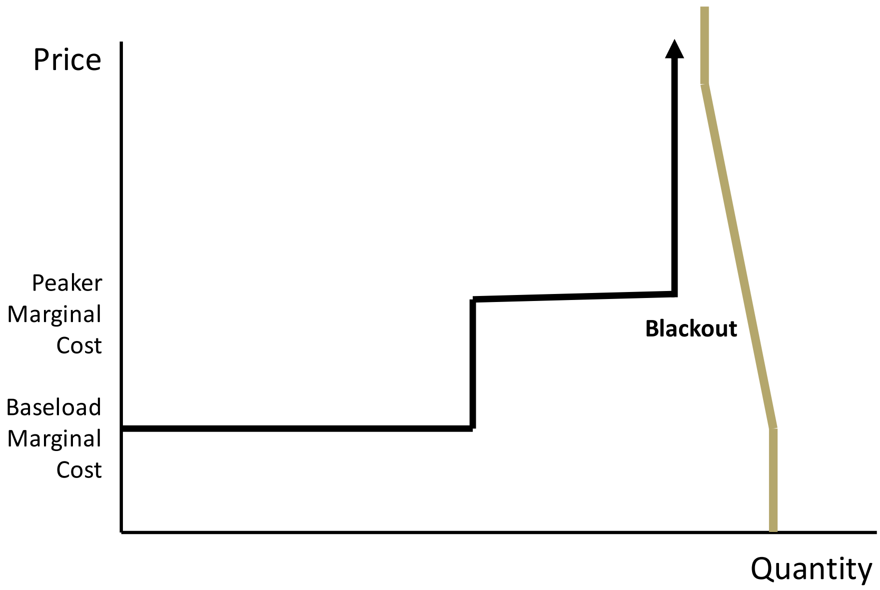

Figure 2.1. Absence of sufficient demand response

However, electricity markets do not reflect this textbook ideal of guaranteed market clearance:[2] A notorious problem of electricity markets is low demand flexibility. One reason is that many consumers do not have the infrastructure like real time meters necessary to see and to respond to spot prices. Instead, most consumers typically pay constant (average) prices, making their demand completely inelastic. Because storage of electricity is costly, the supply side is also inelastic as capacity becomes scarce.[3] As a result, with insufficient demand flexibility, there is a possibility of involuntary load reduction—a blackout—if generation capacity is not adequate (Figure 2.1).

In fact, there appear to be only few observers of the German and other electricity markets that would exclude the possibility of involuntary load shedding due to inadequate generation capacity for the next decade. In this case, the question asked above ("What level of capacity provides optimal reliability?") becomes meaningful. Unfortunately, however, the power market does not provide an answer. The reason is that there are externalities both on the demand and the supply side.

On the demand side, current markets have no access to information concerning how consumers value reliability. In other words, consumers take no market actions that are based on reliability considerations. This is obviously true for consumers that do not have real time meters and who cannot be individually interrupted. These customers have no incentive to cut back when prices are high, because they cannot receive credit for their efforts. They also have no incentive to pay more for reliable service since they typically cannot physically be given more reliability than their neighbor who does not pay for more reliability. The reason is that, with only few exceptions, system operators have no control over the electricity flows that go to individual customers, and can only curtail demand on short notice in relatively large areas composed of many customers. This makes reliability

[2] As we have learned during the crisis, financial markets, too, do not always clear.
[3] But note that there have been substantial advances in pump storage and in the future many will be driving electric cars that offer a large quantity of battery storage when the cars are plugged into the grid.

in current electricity markets a public good, where consumers have no incentive to reveal their 'true' valuation of reliability. On the other hand, this could change with the development of smart grid and smart metering solutions (and possibly faster than capacity markets yield results).

Many large consumers do have real-time meters and some can be interrupted. Also, some consumers may have contracts that specify in advance the spot price level at which the system operator is allowed to implement demand curtailments (sometimes called "priority rationing contracts"). These customers would not have to monitor the spot price in real-time themselves, and still can respond. However, system operators are typically not prepared to black them out based on the performance of their contractual arrangements for power and the performance of their suppliers. But if customers are not blacked out for paying too little, they will not pay for reliability, regardless of whether they have priority rationing contracts. That is, while they may reduce demand if the spot price rises to say €2,000/MWh, this does not reveal the costs they would suffer if they were blacked out. Summing up, the market receives no signal of how load values reliability. Yet without this information, a market cannot determine the optimal reliability level.

As an example, consider an electricity market consisting of two types of generators, baseload and peakers, and hundreds of suppliers with no market power. Peakers are assumed to have a fixed cost of €80,000/MW-year and a variable cost of €100/MWh. There is a demand function that moves left and right at 2 GW per hour with a 4 pm peak of 44 GW on all but 10 hot days when it peaks at 50 GW. Between the prices of €100 and €500 the demand curve is elastic and demand declines by 1 GW. Lack of real-time metering prevents more elasticity, and consumers cannot be disconnected individually in real time. This market is free of almost every flaw normally said to trigger regulatory intervention: There is no meddling regulator, it has no price caps, no offer caps, no market-power (and hence no interventions to mitigate market power), no ancillary services are needed, generation is not lumpy, there is never congestion, demand is 100% predictable, suppliers are free to enter forward contracts, and there is more than the usual demand elasticity. However, the administrator of this market does not know the value of lost load. But then how can the regulator assure that the market that achieves the optimal (efficient) reliability level? The answer is that he cannot (Cramton and Stoft 2006) without the missing infrastructure to mitigate the demand side flaws.[4]

There are also externalities on the supply side of the market (Joskow and Tirole 2007). During a blackout suppliers do not earn money in the blackout area, even though demand is large and capacity is scarce. However, the incentive to provide the reserves needed to avoid blackouts may be too low compared to the incentive in an efficient market system, because all suppliers profit equally from the positive market price resulting from an avoided blackout. The incentive to invest in reserves is particularly small if the price that can be reached in scarcity situations is constrained.

[4] If the administrator knew that consumers have an average value of lost load of €100,000/MWh, the optimal capacity level can be calculated by analyzing a well-designed market with administrative price setting. Whenever the supply and demand curves fail to intersect (a market failure so gross it is not analyzed in any economics text), the price should be set by the administrator to €100,000/MWh to reflect the best interests of consumers. Peakers will earn between €0 and €400/MWh in fixed cost recovery for one hour (half an hour on the way up and half an hour on the way down) on hot days. This will result in €2,000/MW-year of fixed cost recovery out of €80,000 needed. Consequently €7,800/MW per hot day must be recovered from VOLL pricing. This requires 0.078 hours of VOLL per hot day. With demand shifting at 2 GW per hour, demand must exceed supply by 78 MW at the peak. Hence the optimal capacity level is 50,000 MW minus 1,000 MW of demand elasticity minus 78 MW of uncovered demand = 48,922 MW of capacity.

The discussion shows that electricity market prices cannot fully reflect the social costs of involuntary load and supply shedding. Therefore, a pure-market design cannot reveal the information necessary to provide an optimal reliability level. But then, what determines resources adequacy and the degree of reliability in pure-market designs? The answer is: (a) regulatory adjustments in interaction with (b) market power.

(a) Reliability can be the result of (possibly unrecognized) regulatory interventions, such as tampering with the demand for operating reserves. When regulators and system operators are worried about adequacy, they find more reasons to allow prices to rise and are more generous in their demand for operating reserves. More demand for operating reserves implies higher wholesale market prices, which can—in the long-term—induce entry of new capacity. When dispatchers see the system as overly reliable and prices as too high, they may find ways to suppress the spot price. If the reliability problem grows more severe, regulatory intervention becomes more open, in some markets even resulting in the purchase of capacity. However, there is no reason to suppose that any of these adjustments induces the optimal degree of reliability or market efficiency (see also Section 2.4).[5]

(b) Suppliers may be exercising enough market power to attract new entrants, when competitive market prices would still be insufficient to attract new capacity. It is well-known that, for instance, monopoly power can produce more than enough reliability. Some have even argued that electricity markets need market power to allow suppliers to cover fixed costs. While this is wrong,[6] it is true that market power affects market entry and thus reliability. When market power becomes more severe, regulatory and political intervention is likely, and it works to lower prices and decrease reliability. This may sometimes be beneficial and sometimes detrimental. However, there is no reason to suppose that the reliability level due to market power (with or without regulation) is related to the optimal degree of reliability or to market efficiency (see also Section 2.3).

Summing up, given the demand-side flaws, the concept of a market solving the reliability problem without selling a reliability product is logically impossible, until there is sufficient demand response to address the demand-side flaws. Reliability in markets with insufficient demand elasticity is indirectly determined by a subtle interaction of regulatory and market power mechanisms, as we will describe in somewhat more detail below. That is, in current pure-market designs, regulators and administrators implicitly determine the resource and reliability level. As we will show later, a capacity market—if properly designed—explicitly recognizes the need for an administrative solution, but can do better by minimizing market power as well as investment and strategic risks, and at the same time maximizing the scope for competitive forces to determine the price for reliability, the mix of technologies, and the suppliers who build new capacities.

[5] That new capacity was endowed with free CO2 emission certificates reduced net investment costs and thus also incentivized new entry in Europe. Clearly, this too, is not an efficient way to handle reliability issues.
[6] Often, this assertion is based on a misconception of scarcity prices, which do not reflect the incremental cost of the last produced unit, but nevertheless are fully consistent with marginal cost pricing and perfect competition (Section 2.3).

2.2. Spot prices are too low to pay for adequate capacity when capacity is adequate

Demand-side flaws are not the only challenge for resource adequacy. Another problem is "missing money": In 'normal' periods, when there is no shortage of capacity, prices are below the level needed to cover operating and capital costs of new capacity (this subsection), and in scarcity events, prices are unlikely to accurately reflect the scarcity (next subsections).

Figure 2.2. 'Normal' prices are too low to pay for adequate capacity

As an illustration, Figure 2.2 shows a situation where demand is below maximum available capacity. Here, suppliers can earn the marginal production cost of the peakload technology. However, with effective competition, the peakload resources cannot earn costs that go beyond variable operating and opportunity costs, such as in particular capital investment costs (including an appropriate risk adjusted cost of capital).[7] The reason is that investment costs are sunk and so are not taken into account when bidding into a competitive market. In competitive markets, capacity only has a positive price when capacity is scarce (as illustrated in Figure 2.3 below). The crucial implication of this simple observation is that pure-market designs have an inherent tendency to sooner or later produce scarcity: *When generating capacity is adequate*, *electricity prices are too low to pay for adequate capacity*.

With a well-designed capacity market, electricity prices are just sufficient to pay for adequate capacity when capacity is adequate.

[7] That suppliers earn rents on baseload generation is irrelevant for the decision to build peakers. And it is irrelevant for the argument that there is missing money when there is no shortage of resources. In an efficient equilibrium mix of resources, all plants need scarcity prices to cover their total costs. In fact, any type of plant could be used as a benchmark here. See, e.g., Stoft (2002), Joskow (2007), Ockenfels (2007a,b) and Ockenfels et al. (2011) for the pricing mechanisms in competitive electricity markets.

2.3. Spot prices are unlikely to be optimal at scarcity events: Market power

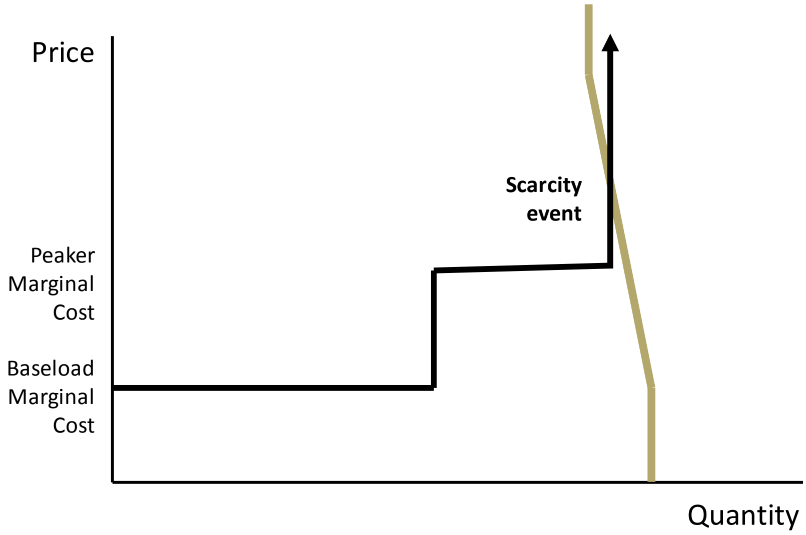

Figure 2.3. Scarcity event

Any pure-market solution needs capacity scarcity to incentivize new resources. If capacity is scarce, there is excess demand at a spot price that is equal to the marginal production cost of the last unit provided by the physically available generating capacity. Because supply cannot do anymore to balance supply and demand in such a scarcity event, the demand side is required to bid prices up such that the market clears. Figure 2.3 illustrates how the resulting market clearing price is given by the intersection of demand and supply at the vertical part of the supply side. The price is called scarcity price and can substantially exceed the marginal cost of the last produced electricity unit. In an efficient equilibrium, the resulting scarcity rents earned in scarcity events are needed to cover the fixed capital and operating costs of all resources. In particular, they are fully consistent with perfect competition and marginal cost pricing.

However, capacity scarcity comes with market power. One reason is that the competitors' ability to substitute others' withheld capacity vanishes as one's own capacity constraints become tight (and because demand elasticity is low, too). In fact, a supplier with even a small percentage of the market could exercise extreme market power when (residual) capacity is scarce. When the system operator needs, say, 96 percent of all generators to meet load, a supplier who owns 5 percent of capacity can exercise extreme market power. Figure 2.4 illustrates in a simplified graph that the scope for profitable capacity withholding market power is extreme in scarcity events (Case I, where prices increase substantially when withholding capacity), while small(er) when capacity is adequate and supply is rather flat (Case II where the same capacity withholding has no price effect).[8]

[8] More detailed discussions and illustrations of the scope for market power depending on capacity reserves can be found, e.g., in Ockenfels (2007b). In fact, unlike in other industries where capacity constraints are less strict due to cheap(er) storage opportunities, market power measures specific to the electricity sector can be read as decreasing functions in excess capacity (DG Competition 2007a,b, Bundeskartellamt 2011). This holds, for instance, for the Pivotal Supplier Index, which shows for every given hour whether a given supplier is "pivotal," i. e. necessary to meet demand. A supplier is more likely to be pivotal when the (residual) capacity level is scarce. Accordingly, market power studies typically found more exercise of market power in peak hours (see Müsgens 2006 for Germany, Borenstein 2002 for the US).

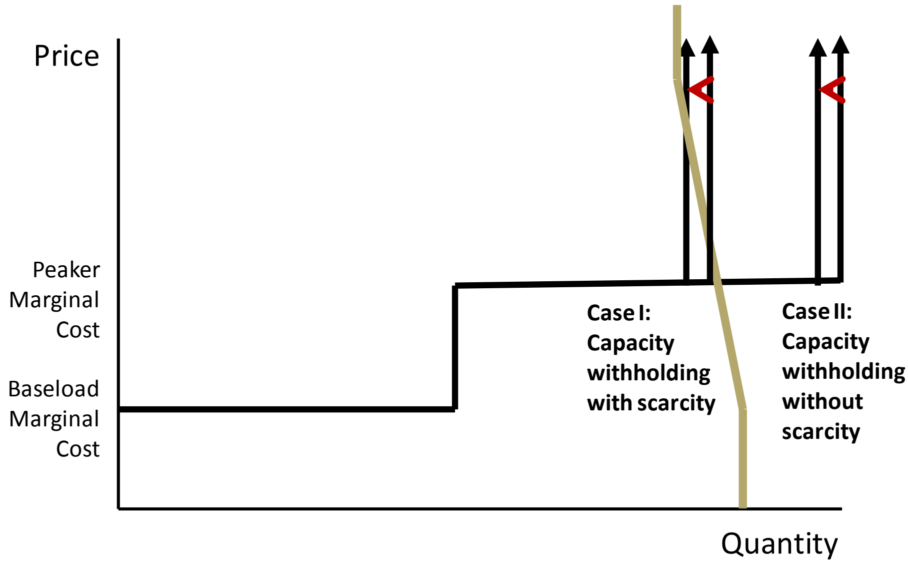

Figure 2.4. Scarcity and market power

Summing up, because scarcity is needed to provide incentives for new investment and because scarcity implies market power potential, any pure-market design will sooner or later run into market power problems—regardless of the market structure. The only way extreme market power in scarcity events can be mitigated in pure-market designs is that market power is already sufficiently large to attract new entry before capacity gets scarce. Obviously, none of these options describes an efficient mechanism to incentivize new entry, or to handle the resource adequacy problem.

As we will see later, a well-designed capacity market, on the other hand, can mitigate both market power and scarcity.

2.4. Spot prices are unlikely to be optimal at scarcity events: Regulation

Market power is likely to trigger regulatory responses. Efficient scarcity prices as illustrated in Figure 2.3 can be very high—say, hundred times higher than the incremental production costs of the last produced MWh. Yet, as we have explained before, the opportunity for suppliers to exercise unilateral market power is also very large when capacity is scarce. It seems difficult, if not impossible, to distinguish efficient scarcity prices from prices reflecting market power during scarcity events.

In fact, there is already confusion in administrations and competition authorities regarding the right definition of competitive prices in scarcity events. For Germany, observe for instance that in-depth analyses of the German electricity market based on many million data points and commissioned by the European Commission (DG Competition 2007a) as well as conducted by the German competition authority (Bundeskartellamt 2011) use incremental costs as the basis of their estimates of hypothetical competitive prices. The definition of incremental costs, however, does not include scarcity prices.[9] If prices never exceeded such incorrectly defined marginal costs, the resulting revenues would be insufficient to attract sufficient entry, and the electricity system would sooner or

[9] See Ockenfels (2008a) for more details and a discussion of this and other flaws in defining and estimating competitive prices in Germany's electricity market. See Nitsche et al. (2010) for one of the very rare attempts to take long-term considerations into account when judging German electricity prices.

later collapse. However, prices that exceed such estimates are threatened to induce regulatory and political measures, possibly deterring new investments.[10] Among the mitigation measures that are discussed or that are already established in Europe or Germany are bid caps, supply obligation measures, market monitoring, and divestiture if prices are deemed to be 'too' high (e.g., Ockenfels 2009c).[11] However, because the condition which leads to high market prices in a well-functioning competitive wholesale market (scarcity) is also the condition when these measures are likely to be applied, such measures might push down market prices below the prices that truly reflect competitive (scarce) supply.

Yet even if we abstract away from regulatory responses to fight market power, there are good arguments to believe that prices in pure-market designs do not lead to an efficient capacity level. The reason is that the price formation process under scarcity conditions, at least in markets like the US, depends critically on decisions made by the system operator (Joskow and Tirole 2007). For instance, before implementing rolling blackouts system operators would often reduce system voltage in order to reduce system demand, stabilizing the system. This, however, has the effect of reducing wholesale prices relative to their 'normal' level and at times when scarcity is strongest. If it then still comes to further non-price rationing (blackouts), the price would effectively fall to zero. This implies that market price signals in scarcity events are biased and will typically lead to underinvestment in reliability.[12] Also, system operators may find it necessary in scarcity events to call on generating capacity with specifically needed properties regarding location and availability to rely on bilateral out-of-market contracts. This too can inefficiently depress market prices received for energy (Joskow 2007 provides examples).

The discussion above implies that there is no reason to suppose that scarcity prices accurately signal the need for new investment. In fact, there are a couple of reasons why wholesale market revenues might turn out to be insufficient to induce efficient entry and to guarantee optimal reliability. This phenomenon, too, is sometimes called "missing money". The reasons are market imperfections, regulatory constraints on prices, and procedures utilized by system operators to deal with operating reserve shortages.

A capacity market, if well-designed, can avoid market power due to scarcity, thus reduce regulatory risks, and protect suppliers against the "missing money" problem induced by system operators, regulators and other administrations.

2.5. Price volatility and risks

There are additional reasons that make capacity markets useful. Electricity markets are characterized by extreme price volatility due to both inelastic and volatile demand and supply. This induces costly risks to both sides of the market. A major determinant of the volatility comes from the need of the market to incentivize entry of new capacity. In fact, a large fraction of the net revenues from sales in

[10] Without long-term contracts or vertical integration opportunism are known to possibly lead to underinvestment, as shown by, e.g., Williamson (1979), Hart (1985), and Joskow (1987).
[11] Price caps that sometimes constrain competitive market prices but overall allow coverage of total costs do not necessarily imply a shortage of capacity, but would induce the wrong mix.
[12] Voltage reductions imply an inefficient use of electric equipment, and so is costly to electricity consumers; these costs, too, are not reflected in market prices.

competitive pure-market designs needed to cover total costs including investment costs is typically generated in a small number of hours when capacity is fully utilized. Yet due to fundamental uncertainty on the demand and supply sides, these hours will erratically fluctuate from year to year (Stoft 2002). For example, electricity markets may experience many 'normal' years with average yearly prices well below average annual costs, implying a long sequence of years with losses for investors. On the other hand, there may be years in which a hot summer or other circumstances lead to extreme price spikes, allowing investors to recover multiple times of their annual fixed costs in each of those years. (The risk faced by consumers, of course, mirror the risk faced by supplier in pure-market designs.)

Due to arbitrage conditions, long-term contracts will reflect average spot market prices, but partly smooth out the year-to-year volatility of the spot market. However, even a fully competitive pure-market long-run equilibrium induces significant random fluctuations in long-term prices caused by runs of 'good and bad luck' in the wholesale market. These are reflected in long-term contracts.

The price spikes due to scarcity events and the corresponding risks on the supply and demand sides may be further magnified, both in the short- and long-term, by various other factors, such as boom-bust investment cycles due to coordination failure (see next subsection), market power and regulatory, political or other interest group's interventions. Possible interventions include adjustments in the emission trading schemes, in nuclear power politics, in antitrust policy, etc., which are mostly decided on a short- or medium-term basis, but significantly affect the annual and long-term profitability of investments in capacity.

All these risks may discourage investors, or require a significant risk premium (see Joskow 2007 for a detailed discussion). As we will argue below, a well-designed capacity market mitigates price risks (we will get back to political risks in Section 4.2). Importantly, the reduction of risk does not come with a reduction of price volatility. Risk is reduced by replacing peak energy rents (the rents derived from selling energy at very high spot prices during periods of scarcity) with a constant capacity payment. In fact, we would argue that the real time price is able to allow greater volatility in times of scarcity, since both load and generators would be hedged. The forward contracting enables spot price signals to work (both politically and financially) and indeed can be as volatile as is required for short-run economic efficiency, as all parties (including load) are exposed to the spot price on the margin (see Section 3 for the details).

Abstracting away from market power issues and risks from exogenous market interventions, Cramton et al. (2006) simulate the risks in an electricity market model. They examined the supplier risk associated with and without a capacity market, which is designed not much unlike what we discuss in Section 3 of this paper. They examined supplier risk using historical market data, as well as detailed market simulations calibrated to the specific market setting. It is shown that in the benchmark case, where demand is assumed to have a constant elasticity of –.05 for prices above the strike price (a 20% increase in price produces a 1% decline in demand), the hedge provided by the capacity market reduces aggregate profit risk (standard deviation of profit) by a factor of 7. More importantly, the hedge reduces supplier risk by a factor of 4.5 in the benchmark case. Even when a high level of demand response is assumed so that prices remain low during scarcity periods and there is less profit risk to start with, the hedge reduces company risk by 55%. Taken together, the simulation results demonstrate the risk reducing benefits of the capacity market.

2.6. Coordination failure

In a pure-market design, the decisions to build new capacity are made independently. This induces *strategic* uncertainty: because one's investment in new capacity tends to be more profitable if others invest less, there are incentives to not or to misinform competitors about one's own intentions. This seems partly reflected by the observation that there is typically a significant gap between the announced plans to build new plants and actually executed plans.

As an example for how strategic uncertainty comes into play, observe that without a capacity market, as the market tightens, it offers an increasingly large prize for the next entrant. However, entry is a secretive process, and so simultaneous entry is possible. Aware of this, investors are torn between holding off until the prize is large enough to support some simultaneous entry and entering quickly to ward off competition. The optimal strategy implies a random element and so the outcome is likely to be inefficient. For instance, strategic uncertainty may result in boom-bust cycles that further increases the price and reliability risks both to investors and consumers. A well-designed capacity market, on the other hand, assures efficient investment coordination without reducing competition.

2.7. The (in)effectiveness of other solutions to deal with such challenges

Capacity markets in the power sector are desirable if they can mitigate market failures that would occur in pure-market designs. Based on the relevant literature and our experiences, we have discussed potential sources of such market failures. We just mention here that there is also empirical evidence, supporting our description of potential market failures, in particular missing money, to provide resource adequacy. Perhaps Joskow (2006, p. 15) has made the most systematic study of this problem:

> "*On the one hand, a market response that leads prices (adjusted for fuel costs) and profits to fall and investment to decline dramatically when there is excess capacity, is just the response that we would be looking for from a competitive market. … At least some of the noise about investment incentives is coming from owners of merchant generating plants who would just like to see higher prices and profits. On the other hand, numerous analyses of the performance of organized energy-only wholesale markets indicate that they do not appear to produce enough net revenues to support investment in new generating capacity in the right places and consistent with the administrative reliability criteria that are still applicable in each region.*"

We have also argued that a capacity market can in principle address these challenges. It can induce just enough investment to maintain adequate resources, induce an efficient mix of resources, reduce market risk, reduce coordination failure, and reduce market power. However, these goals can only be reached if the capacity market is embedded in a sound market and reliable political framework, takes into account the lessons learned from a history of flawed designs, and manages to not get distorted by regulatory and political interests that are not related to the goals of the capacity market. Indeed, a large part of the costs of the adequacy problem in electricity markets has been the distortion of market designs by misguided attempts to solve it. These designs cause risks, inefficiencies and ineffective or even damaging regulatory and political responses (we get back to these points in Section 4).

Before we come to a specific design proposal for a capacity market (Section 3), it appears useful to briefly discuss pure-market alternatives to capacity markets that have been suggested elsewhere.

Increasing demand response. Any energy market does benefit from load that can protect itself against blackouts and market power in scarcity events through increased demand response. Because this moves the market towards an ideal market, the need for a complementary capacity market would be reduced. In fact, it may be better to install real-time meters and use real-time pricing to increase elasticity to the point where the market becomes perfectly reliable with regard to adequacy, than to implement a capacity market. Circumstances could change, for instance, with the prevalence of smart grids and smart metering. This could be quite relevant as (a) it could be faster to create demand response through "smart" approaches rather than seeing the investment effects of capacity markets and (b) it could be cheaper to focus on demand response rather than capacity expansion.

However, as long as demand remains rather inflexible, it cannot fully mitigate market power problems at scarcity events. Moreover, increased demand responsiveness cannot contribute to (fully) solve the missing money, risk and coordination problems. So, even as the pure-market design moves towards an ideal environment, a capacity market might still be beneficial.

One might argue that a capacity market will do harm, because it would typically lead to excess capacity in order to provide reliability. But the costs of excess capacity seem relatively low. The reason is that marginal generation capacity is relatively inexpensive when compared with other costs of delivered energy. In fact, as with most optima, the derivative of net benefit with respect to capacity is zero at the optimal capacity level. For example, an extra 10% of capacity increases capacity costs by much less than 10% because peaking capacity is the cheapest kind of capacity, and adding peak capacity does not increase fuel costs, transmission costs or administrative costs significantly. As a consequence, increasing total capacity by 10% will cost consumers only, perhaps, 2% extra. But there is some benefit to the resulting extra reliability. So the loss of net benefit is less than 2%. A good and unbiased regulatory approach is probably unlikely to overshoot by more than 10% on average, and the best market-based approach will not be perfect. Hence, assuming that regulation does a reasonable job, the net benefit of improved adequacy from any pure-market approach is supposedly quite small.[13]

So, even if the demand side flaws were negligible (they are not), a capacity market might be desirable.

Reserve power markets and capacity payments. It is sometimes stated that capacity markets are not necessary because reserve power markets already provide capacity payments. However, this belief is wrong. Capacity bids in competitive reserve power markets reflect short- or medium-term opportunity costs for withholding capacity from the spot market, plus short- or medium-term (quasi-) fixed costs, minus expected profits from actually being called to deliver reserve energy (e.g., Chao and Wilson 2002, Ockenfels et al. 2011). These payments are *not* related to the costs of new capacity, thus cannot efficiently incentivize new entry, and so cannot efficiently address the capacity adequacy problem. Power reserve markets can only indirectly incentivize investments by influencing the wholesale price: larger reserve requirements would draw away capacity from the wholesale

[13] On the other hand, the costs of a shortage of, say, 2% of capacity is probably orders of magnitude higher. Suppliers, when left to their own devices, would typically prefer to err on the side of underinvestment.

market, increasing the wholesale price and thus creating incentives for investment. Yet, once all adjustments are done, the challenges described above (missing money, coordination, etc.) would be similar.

However, reserve power markets can play an important role to address problems that arise with a strong growth of renewables, as it is the case in Germany. We will come back to this in Section 4.1.

Forward energy-only contracts. It is sometimes claimed that capacity investments must rely on long-term forward contracts for their products as the basis for building production capacity. Resource inadequacy thus may be the result of a lack of hedging—and not because of missing money. However, a host of common examples argue otherwise. Hotels can be more expensive than most generators, yet few rooms are booked before an investor breaks ground on a new Hotel. Volkswagen sells no long-term contracts for cars before it builds the factory to produce them. Chip manufactures build chip fabricating plants before the chips are even designed let alone sold. Long-term investments are rarely financed on the basis of long-term product sales; and long-term investments are often made in industries that sell almost the entire product only a month or two in advance. Why do investors in other industries invest in assets with lives of 10 to 60 years when they have sold none of the output from their investment? Because they expect to sell product at profitable prices. Investments are mostly based on expectations, not on the basis of locked-in forward contracts.

Moreover, forward energy-only contracts do not address the major sources for market failures that we discussed in the previous subsections and that capacity markets are supposed to mitigate. The reason is that such contracts, too, do not deal with some of the critical market imperfections and institutional constraints that may lead to inadequate supply resources. For instance, while hedging contracts mitigate price volatility and market power, they do not affect those spot prices that cause the missing money problem in the first place; because of intertemporal arbitrage, forward prices only reflect these spot prices. Furthermore, forward contracts are medium-term only, and so pricing of these contracts does not take into account long-term (entry) costs. So, unless forward contracting does not incorporate long-term “resource adequacy criteria,” they will not solve the problems described above.

Long-term forward energy-only contracts indeed raise a potential problem to the market. They can be used by large loads to attract excessive resources that cause spot prices to fall excessively, thereby reducing the value of the prior investments of incumbent generators not party to the new contracts.

Similarly forward capacity contracts may play a similar role. They may combine long lead time and certainty for the investor, motivating entry. However, they may occur at levels that suppress spot prices, reducing the value of the prior investments of incumbent generators.

Summing up, what is required to address the challenges discussed in this section is not a pure-market solution, but a ‘mainly-market’ solution. One in which a regulator selects the adequate level of capacity and designs the market parameters to induce the market to provide that level as cheaply as possible. This still leaves the most important role to the market, but the regulator must solve the problem of what level of installed capacity is adequate. The next sections explain how to design such a market and adjust its parameters to solve the adequacy problem while maximizing beneficial and minimizing detrimental side effects. We also discuss some challenges and potential problems.

3. Basic design features of capacity markets

This section describes design principles and design features of an appropriate capacity market. It is based on analyses of various recommendations for market designs for adequate generating capacity, and their convergence (as surveyed by Cramton and Stoft 2006), as well as on experience in actually designing capacity markets such as in Colombia and New England. However, we often do not quantify the parameters, because the quantification ultimately depends on data and information that need to be revealed, discussed and coordinated among the various players and regulators before implementation of the capacity market. If we quantify parameters (such as the strike price, the commitment period, etc.), this is meant as a starting point for the discussion.

3.1. Overview: Essential design elements

As explained before, if there is a reliability problem, then there can be no pure-market solution to it in current electricity markets; it must be solved by regulation or market power. The goal of a new market is then to design a regulatory approach that mimics a competitive market as closely as possible. By setting the spot market bid cap at VoLL and using options to suppress risk and market power while preserving marginal incentives, regulation can be confined to the single task for which it is needed: determining the adequate level of capacity.

We will now describe a simplified design that illustrates the principles involved. Related approaches are described in Bidwell (2005), Chao and Wilson (2004), Cramton and Stoft (2006, 2007, 2008), Oren (2005), and Vazquez et al. (2002). It is assumed to be implemented in a single isolated market. Later we will discuss the challenge of border-crossing trade. At first, the energy market will be assumed to contain only the real-time market, but later, settlements that include forward markets will be described.

Step one of the design is to set the bid cap on the spot market to the best estimate of VoLL. Of course this is a poor estimate, but a value of €3,000 should provide good dispatch incentives. Since VoLL will play no role in determining investment, the level is not critical. The difference in energy supplied under a €3,000 cap and a €30,000 cap is supposedly minuscule, because, with market power well controlled, these prices rarely are reached and by the time the price reaches €3,000, there are not many generators that can provide power and are choosing not to.[14]

Step two of the design is to provide load with a complete hedge in the form of a reliability option with a strike price of say €300. Load purchases the target quantity of physical capacity, established by the regulator, together with a load-following call option at the strike price. In this way, load is 100% hedged from energy prices in excess of €300. If the energy price in an hour is €1,000, each supplier has a financial obligation to serve its share of load. Since deviations are priced at the €1,000 spot price, the supplier continues to be motivated by the spot price, although both load and suppliers are hedged from price volatility above the €300 price.

[14] We note that a bid cap is not necessarily equal to a price cap, because prices could in principle be negotiated outside a market platform that imposes the bid cap. However, this is not critical to our design proposal, because the exact level of the bid cap is not critical, bilateral real-time negotiations are difficult in electricity markets, and because the incentive to withhold capacity in our design are small anyway. What is critical, though, is that—to the extent scarcity is a real-time problem—there is a reliable real-time market price, and that the relevant capacities can be quickly dispatched.

The key point is that all generation still faces the spot price even though it is hedged. Suppose a supplier owns 100 MW of capacity. If it provides 80 MW of power for the hour in question and has a 90 MW obligation, it is paid €80,000 because the spot price is €1,000, but it must pay (90 MW) x € (1000 – 300) as a hedge payment. If it provides 90 MW of power, it is paid €90,000 and is obliged to make the same €63,000 option payment. If it produces 100 MW it is paid €100,000, and again makes the same hedge payment. For every MW it increases or decreases its production, its net revenue increases or decreases by €1,000. Note that when the spot price is €300 or above, it makes sense for virtually every generator to be producing, since marginal cost typically is less than €300. As long as the suppliers produce their share of load, they will earn the strike price for all of their output. In other words, a generator with average performance is nearly fully hedged against spot prices above €300 by its physical generator.

High spot prices send a vast array of accurately targeted performance signals. They do because they reflect, with reasonable accuracy, the actual value of electric power at different times. When the system becomes short of operating reserves, a little more power is worth a lot, and the price reflects that value. When that value signal goes out to all the suppliers, it turns into a signal for every kind of behavior that helps bring more power to the system. These signals extend far beyond actions that can be taken immediately. Because high prices occur predictably, every time the system is short, and because it will be short again in the future, price spikes send signals to prepare for future actions as well as to take immediate actions.

This approach also greatly reduces the risk from weather related price fluctuations. It does not however reduce performance risk. Although all risk is costly and hence undesirable, performance risk cannot be eliminated without eliminating the performance incentive. Because performance risk is also quite small, we decide to ignore it. In the full price market, no one would probably argue for exceptions, even though their excuses would be equally valid (or equally invalid) in a pure-market setting. No matter why a generator is out of service when the price is €10,000/MWh, no one would consider paying it for power it did not produce due to performance risk. Also, because shortages are inherently unpredictable, rewarding those who overcome that unpredictability is one of the benefits of using price as the incentive.

To summarize our progress with the design, the first step (bid cap equals VoLL) assured that the market conforms to a classic competitive design, although it still includes the currently unavoidable demand-side flaws (which lead to the absence of a robust demand response).

Step two (the complete hedge) preserves the dispatch incentives perfectly on the generation side, but destroys the investment incentive, because although generators face the spot price on the margin, their revenues are limited exactly as if there were a price cap at €300/MWh. Why is this progress? Two advantages are obvious: a significant reduction in risk and in market power.

The third advantage is that VoLL has become much less important. It no longer affects investment. If a higher VoLL is used, the bid cap will be higher and the incentive to perform on peak will be greater, but because of the hedge, this will not increase the earnings of generators. This means changing VoLL does not change investment or adequacy. VoLL is only linked to real-time performance and in this role, VoLL makes little difference for the following reason. If VoLL and the bid cap are €2,000, all functioning generators will produce as much as they can, and if VoLL and the cap are €20,000, they will do almost exactly the same thing. Perhaps, at the higher value, some load with real-time pricing will be reduced, or some generator will squeeze out one more megawatt, but little will change. Because the estimation of VoLL is always controversial, this is an advantage.

Step three, the final step of the design, introduces the capacity auction. This sets the payments to generators for providing reliability options just high enough to induce optimal investment and adequate capacity. An annual auction is used to purchase new capacity up to the level required for reliability. These auctions determine the price of reliability options that is just sufficient to induce the required entry. For example, with a strike price of €300/MWh this might result in the average annual loss of €40,000 of revenue per MW of capacity relative to the spot market of step one. This is what we have termed the "missing money." In this case, new entrants will bid the price of reliability options down to €40,000/MW-year. If the cost of constructing new capacity increases or decreases, due to environmental restrictions or new technology, new entrants will bid just enough higher or lower to maintain a normal rate of return.

The result is that the regulator controls the level of capacity, but the market controls the price of capacity and the type and quality of capacity built. Hence the regulatory intervention has been strictly limited to the determination of the one factor about which the market has little information—the adequate level of capacity.

Although the auction design requires care to address the potential exercise of market power, the following simple procedure would work quite well. Each year an auction is held for reliability options, ROs, which take effect just over, say, four or five years in the future.

Existing generators may choose either to enter the auction with a zero bid, or not to sell ROs. New projects are allowed to bid without restriction. The regulator bids a demand curve that intersects the target adequacy level at the most recent RO price and slopes down to the right by 5% in price for each 1% increase in capacity. It slopes up from the same point by 20% for each 1% decrease in capacity. The auction is held using a descending clock procedure (see Ausubel and Cramton 2004). All accepted bids are paid the clearing price, but existing generation receives one-year contracts while new generation may choose any contract length up to seven years. Once a new generator's initial contract expires it becomes an "existing" generator. If no new generation is purchased in a given year, all existing generators that bid, have their contracts extended for one year.

There is one further rule, which assures full hedging and limits market power in both the spot and capacity markets. Any generator that does not sell ROs for its full capacity receives only the spot price capped at the strike price. In other words, such generators in effect provide the hedge without compensation. In fact, a few extremely unreliable generators may opt not to sell capacity, and others will decide not to because they are selling their capacity into another system. These are good reasons and cause no trouble. It is only withholding capacity to exercise market power that is discouraged.

Some designs omit this rule. Objections to it are based on ideological grounds. We find these unpersuasive and prefer to rely on economic analysis. As explained by Hogan and Harvey (2000) during the California crisis, suppliers will not give up their market power for free by entering into long-term contracts. They realize that their market power is valuable, and will extract approximately the value of this market power before relinquishing it by selling ROs. This rule prevents the exercise of market power in the capacity market, as well as the spot market. Only permanent retirements can reduce the quantity of existing supply (see the next subsections for more details). Although mothballing capacity is a legitimate response to certain market conditions, new capacity is not purchased in response.

One question may be why the spot price would ever exceed the option's strike price if all load and generation is fully hedged. Remember that suppliers only hedge a specific quantity of power, their share of load. Power produced beyond this level is unhedged for the supplier. Not only can some suppliers produce more than their share, but some always will, simply because it is impossible for a hundred or more suppliers to all supply exactly their share. In any case, the spot price can be high, depending on market rules, because the system operator bids a high price or because suppliers bid a high price for supply that may be beyond their market share.

The basic design described above is intended only to illustrate the central ideas of a sound capacity market and is missing many important practical features. A number of these are discussed below, but no attempt will be made to present a complete working design, which should in any case be tailored to the particular market.

3.2. Product design

As explained before, the product traded in the capacity market is firm energy: the physically-backed ability to supply energy during scarcity events, where scarcity event is defined by high energy price. Naturally, the capacity market must clear on a frequent (e.g., annual) basis and sufficiently in advance of the commitment period, so that new projects may compete (say, four years) and make the market contestable. The commitment period for existing capacity can be short (e.g., one year), because the investment is sunk and so there is no need for long commitment. Also, shorter commitment reduces risk and the need to arbitrage across years. The commitment period for new capacity must be longer (say, five to twenty years), because longer commitment reduces investment risk, and the capacity price then better reflects cost of new entry. Moreover, new capacity should be allowed to select shorter commitment in qualification. As a result, new and existing capacity are paid the same price in first year of commitment. (The new capacity price is indexed after first year.) Not much unlike what we see for forward energy-only contracts, however, the capacity payment in any given year can differ thereafter, depending on the timing and length of the commitment.

3.3. Auction design

The biggest challenge for market design is to reduce market power on the supply side. Market power is a critical issue, because the costs of existing resources are sunk, and the new resources are only a small fraction of the total resources needed. That is, if market power is not mitigated, larger suppliers owning existing resources could typically unilaterally set the price for new resources.

This subsection sketches several important measures to mitigate market power.

(a) Demand curve

One critical factor for the success of a capacity market is an accurate determination of the demand curve. From an economic perspective, the demand should reflect supply and demand uncertainty, expected political and regulatory constraints, and a safety margin (as well as estimations of the value of lost load if available). It may also reflect the trade-off between building new generation capacity at a given location and building new transmission capacity as a substitute for building new plants in order to increase reliability at a given location.

If there is an annual market, the target capacity is determined for the first year of the commitment period. If there are given transmission constraints, the target must be determined in each zone, along with transmission limits.

In order to be more robust against market power, the demand curve should be made elastic around the target capacity level, as illustrated by the following figure by way of example (see also Subsection 3.1).[15]

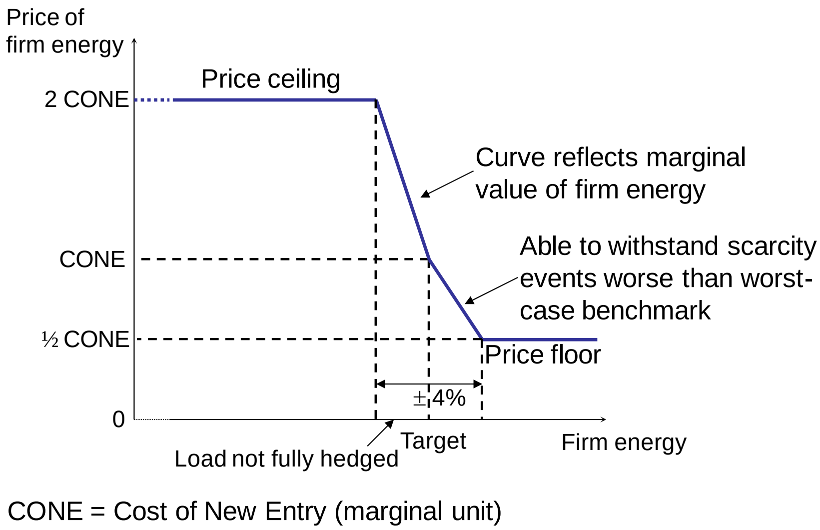

Figure 3.1. Elastic demand in capacity market

We finally note here that establishing the administrative targets must, of course, be set by sophisticated and politically and economically independent planning committees consisting of experienced engineers and economists.

(b) Auction format

We propose an auction to clear the capacity market. It is well-known that the details of auction design can substantially affect the success of a market (Müsgens and Ockenfels 2006). Based on extensive experience in similar environments and well-developed, complementary research (see Ausubel and Cramton 2004 and Cramton and Ockenfels 2011, and the references cited therein), the descending clock auction is the appropriate choice. Here, the auctioneer first announces a high starting price. The starting price must be set sufficiently high (e.g., two times the expected cost of new entry, which in turn needs to be updated for any auction) to create significant excess supply. Then, the excess supply is determined, and the auctioneer announces a lower price.

More specifically, the auction is implemented as a "clock auction" done in discrete rounds. In each round, the auctioneer announces the excess supply at end of prior round, the start of round price

[15] Another possibility would be a vertical demand curve at the target. However, having some elasticity around the target is consistent with what we know about the marginal value of capacity to load and has the additional benefit that it at least slightly mitigates supply side market power.

(higher price), and the end of round price (lower price). Each bidder submits a supply curve in each round at all prices between start of round price and end of round price. The auctioneer then determines the excess supply at end of round price. If there is no excess supply, the auction is over and the clearing price is determined.[16] This auction, along with its restricted feedback policy, minimizes the scope for market power exercise, and at the same time allows useful auction market feedback to facilitate price discovery. Figure 3.2 illustrates the mechanics of the auction.

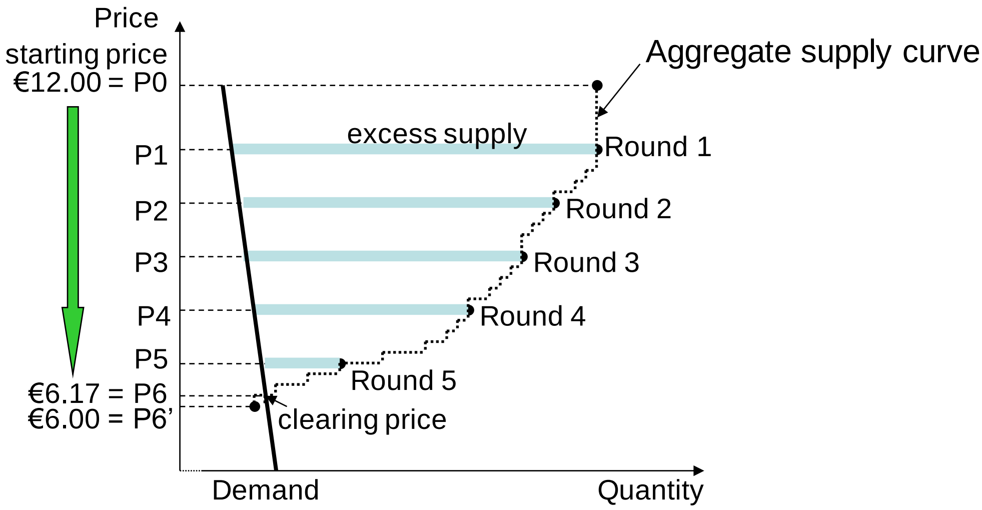

Figure 3.2. Auction design for capacity markets

(c) Bidding and capacity rating

If the auction only buys 2% new capacity in any given year, one cannot depend on enough bids to supply 20% new capacity. No one wants to spend a lot of money planning an investment that has only a small chance of being realized. If the supply curve ends at, say, 8% new entry, then existing suppliers could withhold 6.1% and drive the clearing price to the auction's starting price. The purpose of the capacity market design is to make sure that the market has enough installed capacity, not to induce installed capacity to sell energy. Hence, existing capacity is considered installed capacity, whether or not it chooses to bid in the capacity auction. It can only lose this status by retiring, exporting, or mothballing. To prevent withholding and market power, existing units are counted as supply in the capacity auction unless they request to retire, export or mothball. In these cases, certain restrictions, consistent with the nature of the request, and inhibiting of market power, can be placed on them.

More specifically, to mitigate market power, only new capacity can set the price in the capacity market (this assumes competition for new capacity).[17] That is, bids of new resources are not mitigated. All existing capacity that qualifies for the capacity market, however, is considered to bid in at a price of zero, with exceptions being retirements and imports/exports (see Section 3.5). Retirement bids are submitted before the auction, and accepted retirements are excluded from any

[16] Lumpy investments are respected: the investor does not have to fear partial acceptance. If multiple bidders drop at the clearing price, the group of bids are accepted to balance supply and demand as closely as possible.
[17] This may not be the case if, for instance, there is only a limited number of suitable sites and these are owned by few suppliers. In this case, other measures to mitigate market power are necessary.

future capacity auction in order to minimize strategic behavior.[18] Retirements are replaced with new resources in the auction, represented as a shift to right in the demand curve for all prices below the retirement bid.

Existing capacity can also opt out of capacity market with an exit bid above the clearing price. This might be an attractive option for very unreliable capacities or if suppliers wish to sell capacity into other systems (see also Subsection 3.5).

In the German context, as in other mature markets like PJM and New England, it is important to interpret "new entry" in a broad sense that includes the substantial retrofitting of existing resources to address environmental concerns. Indeed, the marginal capacity price may well be determined from the going forward cost of aging capacity that is required to stay online temporarily before suitable renewable capacity or transmission can be brought online.

It is a necessary condition for these rules that the capacity that qualifies for the capacity market is accurately defined, verified and rated. In particular, suppliers bid a quantity of capacity, but they cannot be allowed to determine the capacity rating of their own units. Instead, a capacity bid should be either its nameplate value or something determined by the administrative engineers. In a system with lots of renewables, this quickly becomes complex. For example, the contribution of wind units to resource adequacy is obviously smaller than the contribution of coal units, and moreover may depend on the capacity mix in the market, location and transmission constraints.

3.4. Absence of a centralized day-ahead market

There have been concerns that the mandatory capacity market requires a centralized day-ahead market such as found in the Northeastern ISO's of the US (PJM, ISO-NE, and NYISO). Fortunately these concerns are unfounded. There is also concern that the RO mechanism is incompatible with a high level of bilateral contracting. In fact there always has been a high level of bilateral contracting in New England and other markets where this design has been adopted. How ROs integrate with bilateral markets and trade with other markets is explained below.

Assume a 100 MW generator has sold a reliability option with a strike price of €300, and the Transmission System Operator, is handling the reliability market. There are also long-term bilateral markets, a day-ahead market run by APX, and a balancing market run by the TSO.

Currently, European markets can be thought of as two settlement systems. A generator sells total energy of $Q_{Forward}$ in forward markets, either bilateral or centralized like APX, and $Q_{Forward}$ is scheduled with the TSO. The generator then delivers an amount Q_{RT} in real time. The forward quantity is settled at the privately determined prices regardless of what is delivered in real time, while the deviation from the forward sale, $Q_{RT} – Q_{Forward}$, is paid the balancing market price, $P_{Balance}$. Of course the deviation can be negative, in which case the generator pays the TSO. So without reliability options or when the balancing price is below the strike price, the settlement works as follows:

Generator Revenue = $P_{Forward} \times Q_{Forward} + P_{Balance} \times (Q_{RT} – Q_{Forward})$

[18] Retirements may be rejected for reliability reasons, but only if the reliability problem cannot be resolved during the planning period with alternative actions, such as transmission upgrades or new capacity.

With reliability options, and the balancing price above the strike price, the RO is settled second in what is essentially a three-settlement system. Each supplier is responsible for a share of the real-time load that is proportional to the quantity of reliability options it has sold. For example if it has sold 100 MW of options out of a total of 10,000 MW of reliability options, it is responsible for 1% of the load in every hour.[19] Call the supplier's share Q_{Share}—its share in MW. The three-settlement system works as follows:

Generator Revenue = $P_{Forward} \times Q_{Forward} + P_{Strike} \times (Q_{Share} - Q_{Forward}) + P_{Balance} \times (Q_{RT} - Q_{Share})$

If the generator supplies exactly its load share, so that $Q_{RT} = Q_{Share}$, then it is fully hedged against the balancing price. However, if it deviates either up or down from its share, it is paid or must pay the balancing price. Hence its incentive to perform has not changed, and the balancing market continues to play its traditional role. As can be seen, there is not much interaction between the forward transactions and the reliability option. In particular there is no reason generators cannot sell all of their power in the forward markets and sell reliability options for all of their capacity in the capacity market. Three things have changed as a result of the hedge built into the reliability option: the average generator earns at most the strike price in the balancing market, both load and generation are less at risk, and the forward contract needs only to cover prices below the strike price.

Notice that the sum over all generators of Q_{RT}, the total power delivered, equals the total load, which equals the sum over all generators of Q_{Share}. This means the sum over all generators of $P_{Balance} \times (Q_{RT} - Q_{Share})$ is zero. This is exactly true, because generation shares are determined after the fact. These terms are the incentive payments for under and over performance by generators, relative to Q_{Share}. So these payments do not affect load, but are payments from poorly performing generators to the better performers. As a consequence of this fact, generators do not have to be concerned that the TSO will under-procure capacity because, on average, this cannot change the total performance payments to generators which always sum to zero. Under-procurement of capacity will only increase the number of hours when there is a shortage and generators are paid the strike price.

3.5. Trading between different zones and other concerns

In an interconnected electricity market the actions of any element in the system impacts all others. For instance, the operation of German power reserve markets may affect French electricity prices. However, the interaction of different markets in different zones and for different products such as electricity and reliability options does not necessarily hamper (inter-market) efficiency. In fact, in this section we argue that implementing a well-designed capacity market in Germany does not threaten the functioning of the European cross-border market. Cross-border trade may, to the contrary, often reduce costs and increase efficiency. For one, cross-border trading in both the electricity and capacity market typically reduces the costs of 'German' reliability. Suppose, for instance, that French capacity

[19] Why base the hedge on load share? New England has about 30 GW of capacity, but sometimes, due to cold weather, many generators cannot run, and the price has spiked with as little as 20 GW of load. If reliability options covered the full 30 GW, then load would be paid for 30 GW times the \$1000 spot price less the \$300 strike price. Hence load would profit by \$7 million dollars per hour during such an incident. This upsets generators without reason, and causes them to worry that extra capacity will be purchased so load can profit more in this way. Basing the reliability option on load share solves this problem by putting the generators in a nearly balanced position in every hour.

bids into a German capacity auction. Observe that, according to our product definition, French capacity can only participate in the German capacity auction to the extent it can actually deliver 'reliability' in Germany in times of scarcity in Germany. Then, if French capacity is built because it successfully participated in Germany's capacity market, Germany benefits from lower costs of reliability (compared to a situation in which only German plants are allowed to bid into the market). Moreover, both Germany and France may benefit from generally lower spot prices due to more capacity in the cross-border market (those suppliers successful in the capacity auction get compensated). Also, cross-border trade can make new investments more profitable, which further decreases reliability costs in Germany's capacity market.

The examples also illustrate that a German capacity market may exert positive external effects on France's load. In fact, market efficiency implies that any resource built in Germany, with or without the help of a capacity market, can at certain times potentially reduce French electricity prices and vice versa, and thus exerts positive external effects. Moreover, if the two markets were fully integrated, reliability would be a public good; there cannot be reliability in one market without reliability in the other market. Thus, in this specific case of fully integrated markets, we strongly recommend to align the design and implementation of a capacity market.

That said, we emphasize that a well-designed capacity market is the efficient (but probably not fair) way of guaranteeing resource adequacy (section 2), even if one fully integrated market does not participate in the costs. Moreover, if markets are *not fully* integrated, such that transmission constraints bind during periods of scarcity in Germany, reliability in Germany becomes a private good. In this case, cross-border trade does not require a joint capacity market. In fact, implementing a capacity market that spans *distinct* markets but produces a *single* capacity price would be inefficient and ineffective.

Another potential concern is if the price is high in Germany, but, say, Dutch generators are under reliability options, they might leave the Dutch market where, on average, they can earn only the strike price, and sell as much as they can into the German market. To examine this possibility, the settlement must include a term for exports. Since the concern is with the effect of reliability options, the real-time price must be above the strike price, so it is safe to assume a generator (who is counterparty to a reliability option) will produce the most it is capable of, Q_{Max}. Next note that only the power delivered to the domestic balancing market, $Q_{Domestic}$, receives the balancing price. The export quantity is then, $Q_{Max} – Q_{Domestic}$, and the settlement works as follows.

$$\text{Generator Revenue} = P_{Forward} \times Q_{Forward} + P_{Strike} \times (Q_{Share} - Q_{Forward})$$

$$+ P_{Balance} \times (Q_{Domestic} - Q_{Share}) + P_{Export} \times (Q_{Max} - Q_{Domestic})$$

Because the derivative of Generator Revenue with respect to $Q_{Domestic}$ is $(P_{Balance} - P_{Export})$, the incentive to export is exactly the same as without reliability options. Hence there is no justifiable concern with a disruption of the balance of electricity trade, or a collapse of the domestic market.

Another practical concern is the assignment to load serving entities (LSEs) of responsibility for the cost of ROs. Because the options are procured by the TSO, the LSEs are not burdened with purchasing ROs and need not make any long-term commitments by purchasing them. This is a great advantage because it means LSEs are at little risk from consumers moving from one LSE to another. The cost assignment is simply adjusted each year and based on the LSE's coincident peak load during the year. This determination is best made after the fact. For example, option cost responsibility for 2010 should be based on the peak loads during 2010. To reduce randomness in loads on any given day, a weighted average of the three highest peak-load days could be used.

Another practical concern is that load may not be exposed to the spot price because of the hedge. This concern is addressed by hedging load's expected energy demand, rather than its realized demand. For this it is necessary for customers to have real-time meters, as is becoming more prevalent. Moreover, LSEs can implement various programs to encourage conservation during times of peak load, and it would be worthwhile to properly motivate them to do so. This can be accomplished as follows. First compute the peak energy costs of each LSE. This is simply the integral of its load times Max(0, ($P_{Balance}$ – P_{Strike})) over the year. Then, since each LSE is assigned a reliability share, L_{Share}, based on its coincident peak load, this share can be used to compute its share of the total of all peak energy costs. Each LSE then pays a penalty equal to the amount by which its actual peak energy cost exceeds its share of total peak energy costs.

Peak-Load Penalty = peak energy cost – L_{Share} × (Sum of all peak energy costs)

The sum of the penalties is zero, and the derivative of a load's penalty with respect to its own peak energy cost is (1 – L_{Share}), which is near one as long as LSEs are small. This means that purchasing a MW of power, when the balancing price is above the strike price, costs each LSE an amount P_{Strike} + (1 – L_{Share}) × ($P_{Balance}$ – P_{Strike}), which is very nearly $P_{Balance}$. In other words the penalty makes the LSEs face the balancing price on the margin even though they pay no more on average. There will be some risk to loads from this performance penalty but it is small and it is only what is inevitable if loads are to face the real-time price on the margin. They are still completely hedged against price spikes caused by weather, nuclear outages, or other events out of their control. This same technique can be used by LSEs to pass real-time price signals through to their loads equipped with real-time meters.

One final concern is that reliability options may impose burdensome new information requirements. However, the TSOs are already aware of the quantities transacted in the forward markets because these must be scheduled. They are also aware of actual production and of which generators are exporting how much power. It also knows the daily loads of the LSEs, and the balancing market prices. This is all the information needed to implement this reliability option design. In particular there is no need to collect more information about bilateral transactions.

3.6. Demand side integration

Qualified demand can provide reliability. In fact, we recommend to fully integrate the demand side in a way that is completely symmetric to the supply side; that is, qualified demand should compete directly with resources on the supply side such that curtailments must be reliably provided, contingent on the strike price in the spot market. Demand is paid the same capacity price for the same reliability option as supply.

However, often the demand side may not supply the necessary flexibility and reliability as required for capacity markets. Yet, demand can then contribute to reliability in power reserve markets, and there too should be treated symmetrically to the supply side. Capacity markets seem especially attractive to the demand side if some non-trivial investment costs need to be sunk in order to provide reliability options.

To conclude this section, we note that our list of design features is not exhaustive. Other design issues include specific rule for the transition phase (such as ceiling and floor on firm energy payment to existing suppliers), procedures for insufficient supply and insufficient competition events, and penalties for not delivering reliability. While the details are important, a detailed discussion of such features would go beyond this paper, partly because they will depend on other details which need to be settled first.

4. Other issues affecting resource adequacy and the effectiveness of a capacity market

This section discusses three important issues in the broader environment of Europe's electricity market that have been argued to potentially affect the effectiveness of a capacity market.

4.1. Growth of renewables' capacity

The electricity market in Europe, and in particular in Germany, is characterized by a politically-induced strong growth of renewables capacity, in particular solar and wind energy. The renewables are paid a subsidy for the electricity provided, which is independent of the electricity price. A major economic effect on conventional, market price-driven generation is that residual demand and thus 'normal' price levels decrease, and that price volatility increases. At the same time, however, conventional capacity must exit the market at a much slower rate than renewables enter, because sometimes the sun does not shine and the wind does not blow. As a result, the degree of capacity utilization of conventional generation is significantly reduced. Taken together, all these effects imply that the 'missing money' problem is becoming more severe as the renewables' share grows. A capacity market as described above can help to mitigate the various challenges for conventional generation in a system with a large share of renewables.

However, in times of a *transition* to a system with a large share of renewables, in which renewables are *added* on a large scale to an existing set of power plants, there is probably excess capacity, and so there is no need to incentivize new entry through a capacity market. Rather the challenge in the transition phase might be to incentivize efficient conventional generation capacity to *not* exit the market and to instead provide reserves. In fact, some observers argue that the increasing share of renewables requires additional payments, beyond spot prices, to old, existing plants. One argument is that, in the short-run, a strong growth of renewables leads to lower prices, because they create excess capacity, and thus cause stranded costs (as may have happened in Spain). While the argument is correct, it does not imply capacity payments from an economic perspective; whether suppliers should be compensated for such stranded costs is a distributional and not an economic question. In the long run, excess capacity is reduced and then the question arises whether there is sufficient incentive for new capacity to enter the market. Whether there will be resource adequacy will then depend on the factors that we discussed in Section 2, which can be mitigated by the market described in Section 3.

However, some argue that the strong growth of renewables, together with large political uncertainties (see next section), requires complementary measures in order to avoid *economically inefficient* market exit of old plants, which are about to be decommissioned. It is stated, for instance, that old hard-coal plants may be increasingly needed as a reserve for renewables (and, more generally, to fill the sometimes predicted 'electricity gap') but cannot cover quasi-fix costs in the wholesale market because of decreasing utilization and missing money due to regulatory interventions, and because of too little flexibility to bid in reserve markets. For this reason, some argue that incentives such as capacity payments or a 'strategic mothball reserve' operated, owned or leased by an administrator is needed to keep the old units in the system (see, for example, the case of Sweden; Frontiers 2011).

We caution that such instruments risk to unnecessarily reduce market efficiency. For instance, in Sweden the 'strategic reserve' is dispatched at a price determined by last commercial bid. This will likely suppress new investments and distort prices away from efficient spot market pricing.[20] That is, this strategic reserve jeopardizes both the short- and long-run efficiency of the electricity market. It also cannot mitigate market power and price risks that come with capacity scarcity. Also, it cannot solve the missing money problems discussed in Section 2, because—if at all—the system tends to further suppress spot market prices. It is thus likely that this system will harm market efficiency and long-term reliability.

A concern of insufficient and inefficient incentives to stay in the market can be mitigated without such distortions. First, such insufficient incentives to stay in the market often come from insufficient transmission or a lack of locational pricing. For example, in Germany, many observers would probably agree that there is a risk of insufficient capacity in the south, but not in the north. The actual source of such geographical discrepancies is insufficient incentives to build generation and/or transmission capacity where it is needed most. However, locational concerns in the spot market must be resolved either with locational marginal pricing or ample transmission for a capacity market to prove effective. Otherwise, generation investments will inevitably be built in the wrong locations, and lead to inadequate resources.

Second, another source for insufficient incentives to stay in the market is an inappropriate design of the reserve power market. If well-designed, power reserve markets make sure that flexible-enough generators (such as gas plants) can earn economic rents sufficient not only to cover incremental production costs but also to cover quasi-fixed costs (short-run fixed costs). This holds even if the existing capacity is only rarely called to actually deliver electricity. One key challenge in Germany's transition phase is to adjust the specifications of reserve levels and products in an increasingly renewable-dominated system. If, for instance, efficient balancing of a renewable-dominated system requires also less flexible resources to stay in the market (such as hard coal plants, which cannot quickly ramped up), an efficient reserve power market must adjust by adding or redefining the product portfolio accordingly (e.g., reserve that is less flexible but available for longer time-periods). An appropriate power reserve market design, one that reflects the reserve requirements of the system, implies that suppliers will keep their resources in the market as long as they can provide these products at lower costs than new entry. Because power reserve products are interlinked with each other through arbitrage, competitive reserve markets with appropriate product definitions will lead to an efficient, cost-minimal provision of reserve capacities in the short- and medium-term.[21] Together with a well-designed capacity market, a well-designed power reserve market will then lead to long-term efficient exit and entry decisions.

[20] The efficient price would be the value of lost load, which is however not known and if it were know, it would probably not be politically acceptable. In fact, to our knowledge, no such system even attempts to employ value of lost load bidding.
[21] An increasing share of renewables also requires an integration of intraday and balancing markets.

Finally, we note that prices paid to suppliers in our capacity market as sketched in Section 3 cover any missing money both, for existing and new capacity. Because old plants commit only from year to year, old plants compete from year to year with new plants to provide reliability. If they decide to drop out of the market, this is because new capacity is more efficient. Moreover, a variation of our capacity market can, in principle, also induce efficient *exit* (and not only entry) of capacity. Here, the capacity auction would determine those plants that would need to exit the market, and the price paid to all capacity staying in the market would be set by the going-forward cost of the last exited plant. (There are several design parameters and challenges that are somewhat different than what we discussed above, though.) However, such a model should only be complementarily considered, once transmission, locational pricing mechanisms, and power reserve market products have been optimized.

4.2. Politically-induced uncertainties

One of the major risks faced by investors in electricity markets comes from political uncertainties. One notorious example is the remaining operating time of existing nuclear plants in Germany. On 30 May 2011, Germany committed to shutting down eight nuclear plants immediately and the remaining nine by 2022; only the three most modern nuclear plants will be allowed to produce electricity at the 2022 final deadline. Political decisions are made, withdrawn and then reversed, depending on which political party is in charge, on court decisions, and on evaluation of an ethics committee. Obviously, this creates substantial uncertainties regarding the demand for new capacity and the profitability of investments—regardless of whether there is a capacity market or not.

Another example for excessive politically-induced risk are moving renewables and CO2 targets, as well as suboptimal instruments to reach those targets. For instance, the price-control ('feed-in-tariff') mechanism employed to subsidize renewables leads to large uncertainties about the growth of renewables and thus about the net demand for electricity.

Finally, there is also the risk of political and regulatory opportunism, when prices reach levels that are economically sound, but judged not acceptable (possibly based on confusion about what should be expected in a competitive market; Section 2.3).

Even without any risk aversion, such uncertainties can lead to significant delays of new entry, because the investors will wait for further information that specifies the project's future cash flows ('real option theory'). A capacity market can in principle mitigate the timing problems, because the entry timing is determined through the auction. However, a capacity market cannot avoid drastic risk premiums that need to be paid to investors in hugely uncertain environments, or even unwillingness to invest at all in unpredictable environments. That is, with or without capacity markets, large politically-induced uncertainty makes reliability of electricity supply an impossible or expensive goal. In particular, observe that any plant that provides reliability options, including the existing plants, would be paid large risk premiums, making reliability very expensive in times of profound political uncertainties.

So, complementary to considering a capacity market, one natural approach to address resource adequacy problems is to implement a stable political framework that takes into account its long-term effects on the electricity sector. In fact, the immediate benefits from a predictable market framework for reliable electricity supply may well be larger than the immediate benefits from a capacity market. There are many ways to make capacity demand more predictable. To give only one example, an auction that uses a quantity- based approach to subsidize, say, solar energy, instead of a feed-in tariff, would allow to fully control timing and quantity of new solar capacity, and at the same time minimize the costs of the target.

4.3. Regulatory imperfections

We argued that a well-designed capacity market can coordinate new entry in ways that make electricity markets more robust against demand side flaws, market power, uncertainty, etc. However, it is increasingly recognized that market imperfections on the one hand need to be balanced against regulatory imperfections on the other hand. Mistakes in the 'engineering' of financial markets (Bernanke 2010), spectrum markets (Cramton and Ockenfels 2011), electricity markets (Cramton 2003, Müsgens and Ockenfels 2006) and other sectors can turn a good design idea into a fiasco. An example is California's energy market in summer 2000, which is widely regarded as a market *design* failure—not a market failure—and which illustrates that ignoring the economics of energy markets may lead to catastrophic outcomes.

This also holds for capacity markets. In Germany, the mix and geography of (new) capacity is subject to strong pressures from market, political, environmental and other lobby interest groups. [22] It is likely that such interest groups will try to bias the design—e.g., the ratings of certain technologies and locations—towards their interests. On the other hand, however, one might hope that a transparent debate about how much capacity is actually needed may help to discipline the influence of interest groups. We illustrate this trade-off by way of example.

A roughly accurate prediction of electricity demand is a key condition for a capacity market's success. But will administrators come up with reasonable estimates? The German government recently announced ambitious energy efficiency goals leading to decreasing electricity demand in the next years. However, many observers expressed concerns that without strong and expensive improvements in the efficient use of electricity, the government's climate goals appear unlikely to be reached, and that, in fact, demand may to the contrary further and significantly increase.

---

[22] Based on his long experience in regulating the US electricity and other sectors, Joskow (2010) describes what we think is a not an atypical phenomenon in the administration of market regulation and design: "The regulatory process is subject to interest group capture, political influence, and tremendous pressure to engage in (hidden) taxation by regulation (Stigler 1971; Posner 1971; Noll 1989). The modern field of political economy based on rational actor models of political behavior did not start with studies of regulation by accident. This phenomenon goes well beyond simplistic models of capture by regulated firms and reflects the fact that regulatory agencies have things that they can do to help one interest group and harm others, naturally leading them to become targets of political competition. This phenomenon is exacerbated over time as young 'expert' regulatory agencies become dominated by commissioners and senior staff who have come up through the political process and are sensitive to the same political considerations as are their sponsors in the executive and legislative branches and those they regulate. In my view, this has become a more serious problem over time as 'independent' regulatory agencies once heavily populated by reasonably independent technocratic experts with clear public interest goals have increasingly come to be populated by commissioners and senior staff with narrower political goals—whether it is on the less regulation or more regulation extremes of the political spectrum depending on which political faction is in power."

A capacity market forces the administrator to procure capacity based on judgments about expected load. If such judgments get distorted in order to maintain politically desirable illusions, the ability of the electricity market to reach efficient outcomes will suffer. On the other hand, however, if such judgments are not just cheap talk and discussed in a transparent way, policymakers may refrain from creating illusions.

Similarly, capacity markets forces administrators to explicitly deal with the costs, uncertainties, and capacity implications induced by time-inconsistent nuclear power politics, inefficient subsidies of renewables, local resistance to new conventional power plants, etc. On the one hand, capacity markets make the implications of such politics for investment costs transparent. This can lead to more rational debates and decisions. On the other hand, there is a danger that the parameters in the capacity market get distorted because of pressure from politics and lobbyists, or because the capacity level is used as a strategic commitment device, to constrain the decisions of future policymakers and investors.

Regarding competition policy, there are concrete ideas that small utilities should get subsidies when building new capacity. While such policies could in principle be embedded into a capacity market, capacity markets already mitigate investment risks and market power *per se*.

There is a history of flawed capacity market design, resulting in large inefficiencies and costs. Thus, if a capacity market is going to be established, we recommend installing an *independent* group of experts to manage and supervise the process.

5. Discussion and conclusion

The forward reliability market approach described here is the product of a systematic development based on clear economic principles. Indeed, once it is understood it seems almost inevitable in order to deal with the resource adequacy problems described in Section 2, and as described in the literature.

An important step is to suppress risk and market power by introducing reliability options. These do not interfere with real-time price signals, so the market retains the dispatch optimality of classic market pricing, but without two major problems. Yet the second step also suppresses a significant amount of generation revenue and thereby destroys investment incentives. The capacity market, however, induces the adequate level of investment by procuring the appropriate quantity of reliability options. This is something the market cannot do because of demand-side flaws. All generation, new and existing will want to sell reliability options for their full capacity because these options fetch a high price relative to the financial cost of the option. Participation in this market is guaranteed by the rule that non-participating generators receive the spot price capped at the option strike price.

The auction for procuring reliability options takes place several years in advance of the effective date, so that there is time for new entry to back the options. To suppress market power in the reliability-option auction, only new capacity bids are allowed to set the price. The auction is a descending clock auction. By inducing investment with an auction (instead of high prices), not only does the regulator have better control of the average reliability level, but far better investment coordination is assured.

While there are many concerns about the use of reliability options, with the designs specified here, many are not warranted. For instance, there is no difficulty in deciding which private contracts are acceptable substitutes for reliability options, because no substitutes are accepted. One hundred percent coverage by reliability options does not interfere with 100% coverage with bilateral contracts. Reliability options provide price coverage above the strike price; bilateral contracts provide price coverage below the strike price. Although reliability options limit the average real-time price to the strike price, the marginal price for both load and investors remains the balancing market prices. This preserves incentives and prevents any increase in exports relative to the present system even on days when other countries have high prices.

The benefits of this design, regarding the resource adequacy problem are significant. It allows:

- explicit and transparent control of resource adequacy
- coordinated new entry (no boom-bust cycles, no strategic delay due to political uncertainties etc.)
- minimum cost of new capacity (through competitive bidding)
- reduced risk premiums (savings to load; stable capacity price does not imply that fixed-cost recovery is guaranteed: a supplier with higher fixed costs gets no mire and one that fails to perform loses its entire capacity payment)
- reduced market power in the spot market (supply cannot profit from price spikes, including price spikes from withholding)
- ideal performance incentive (suppliers feel the full performance effect of high spot prices)

That said, we acknowledge that the current public discussions surrounding capacity markets in Germany and Europe are partly based on different problems that cause reliability concerns than the one discussed in this paper. Reliability concerns in Germany seem mostly rooted in the transition process of Germany's electricity market, away from nuclear power and towards renewables. This transition comes with strong and hard-to-predict growth of renewables, time-inconsistent and drastic interventions of energy politics and regulation, ambiguous development of the grid, and a concern that existing capacity may exit the market and does not serve as power reserve. The reason we only marginally speak to these challenges is that capacity markets *cannot* be part of the solution to these problems. No capacity can mitigate the risks implied by, say, time-inconsistent politics (although it can mitigate the large price risks that are already inherent in stable markets). Rather, a capacity market makes the corresponding costs for long-term reliability transparent. Moreover, economics would require that, with a capacity market in place, all capacity that provides reliability should be paid the same price. This implies that the risk premiums due to an unstable and unpredictable electricity market framework is paid to all capacity, including the existing capacity. The economic cost of reliability is thus likely to be large due to the currently unstable political and regulatory framework.[23] Finally, in times of rather emotional public discussions, ambitious (or even illusionary) goals on future technologies and demand, and a financially strapped electricity sector, it may seem unlikely that the market design will be based on rational economic arguments only; yet flawed designs likely do more harm than good.

We thus recommend in the current phase of transition in Germany, to give highest priority to build a stable and reliable political, and a sound market framework. This should include controlling the path of entry of renewables, integrating renewables into the wholesale and reserve markets, credibly committing on nuclear power, CO2 and renewables politics, and new transmission lines, adjusting the design of power reserve products and markets, reduce regulatory uncertainties, strengthen real time demand response, real time price discovery, and real time dispatch flexibilities, as well as other measures. While the details of these measures go beyond the scope of this paper (yet are discussed in the literature), most of them could substantially contribute to resource adequacy and reliability. In fact, given the current state of Germany's electricity market, the contribution from building a stable and more flexible market environment will likely exceed any contribution to reliability from well-designed capacity markets. That said, a capacity market, as sketched in this paper, will be of important complementary value once the market operates on a stable political, regulatory and economic basis.

6. References

Allaz, B. and J.L. Vila (1993), "Cournot Competition, Forward Markets and Efficiency," *Journal of Economic Theory*, 59: 1-16.

Ausubel, Lawrence M. and Peter Cramton (2004), "Auctioning Many Divisible Goods," *Journal of the European Economic Association*, 2, 480-493.

Bidwell, Miles (2005), "Reliability Options," *Electricity Journal*, June, 11-25.

Boiteux, M. (1960), "Peak Load Pricing," *Journal of Business*, 33:157-79 [translated from the original in French published in 1951.]

Boiteux, M. (1964), "The Choice of Plant and Equipment for the Production of Electric Energy," in James Nelson, ed. *Marginal Cost Pricing in Practice*, Englewood Cliffs, N.J., Prentice-Hall.

Borenstein, S. (2002). "The Trouble with Electricity Markets: Understanding California's Restructuring Disaster" *Journal of Economic Perspectives*, 16(1), 191-211.

Borenstein, S., J. Bushnell und F. Wolak (2002). "Measuring Market Inefficiencies in California's Wholesale Electricity Industry" *American Economic Review*, 92(5), 1376-1405.

Bundeskartellamt (2011). "Sektoruntersuchung – Stromerzeugung – Stromgroßhandel," available at http://www.bundeskartellamt.de/wDeutsch/download/pdf/Stellungnahmen/110113_Bericht_SU_Strom__2_.pdf

[23] Moreover, the details of the auction and product design of capacity markets (e.g., with respect to the rating of capacity bids) with unclear future transmission constraints and technology mixes are likely to be much more complicated than what has been implemented before in other, more settled markets.

Bushnell, James (2005), "Electricity Resource Adequacy: Matching Policies with Goals," CSEM Working Paper No. 146, August. http://www.ucei.berkeley.edu/PDF/csemwp146.pdf

Chao, H.P. and R. Wilson (1987), "Priority Service: Pricing, Investment and Market Organization," *American Economic Review*, 77: 89-116.

Chao, H.P. and R. Wilson (2002). "Multi-Dimensional Procurement Auctions for Power Reserves: Robust Incentive-Compatible Scoring and Settlement Rules," *Journal of Regulatory Economics*, 22(2), 161-183.

Chao, H.P. and R. Wilson (2004), "Resource Adequacy and Market Power Mitigation via Option Contracts," EPRI, Palo Alto, CA.

Cramton, P. (2003). "Electricity Market Design: The Good, the Bad, and the Ugly," *Proceedings of the Hawaii International Conference on System Sciences*.

Cramton, P. (2004). "Competitive Bidding Behavior in Uniform-Price Auction Markets" Proceedings of the Hawaii International Conference on System Sciences.

Cramton, P. (2006), "New England's Forward Capacity Auction," Working Paper, University of Maryland.

Cramton, P., and A. Ockenfels (2011). "Ökonomik und Design von Kapazitätsmärkten im Stromsektor." Energiewirtschaftliche Tagesfragen, 61.Jg., Heft 9: 14-15.

Cramton P., and A. Ockenfels (2011). "Germany's 4G Spectrum Auction: Design and Behavior," mimeo.

Cramton P., Y. Shoham und R. Steinberg (2006). *Combinatorial Auctions*, Cambridge: MIT Press.

Cramton, Peter and Steven Stoft (2005). "A Capacity Market that Makes Sense" *Electricity Journal*, 18, 43-54.

Cramton, Peter and Steven Stoft (2006), "The Convergence of Market Designs for Adequate Generating Capacity," white paper for the California Electricity Oversight Board, March 2006.

Cramton, Peter, Steven Stoft, and Jeffrey West (2006), "Simulation of the Colombian Firm Energy Market," Working Paper, University of Maryland.

Cramton, Peter and Steven Stoft (2007), "Colombia Firm Energy Market," *Proceedings of the Hawaii International Conference on System Sciences*.

Cramton, Peter und Steven Stoft (2008). "Forward Reliability Markets: Less Risk, Less Market Power, More Efficiency" *Utilities Policy*, 16, 194-201.

Crew, Michael A. and Paul R. Kleinforder (1976), "Peak Load Pricing with Diverse Technology," *Bell Journal of Economics*, 7(1): 207-231.

DG Competition (2007a). "Structure and Performance of Six European Wholesale Electricity Markets in 2003, 2004 and 2005." DG Comp 26th February 2007, Prepared by London Economics in association with Global Energy Decisions.

DG Competition (2007b). "Energy Sector Inquiry der Europäischen Kommission." (2007). Available at: http://ec.europa.eu/comm/competition/antitrust/others/sector_inquiries/energy/#final

Grimm, V., A. Ockenfels and G. Zoettl (2008a). "Strommarktdesign: Zur Ausgestaltung der Auktionsregeln an der EEX" Zeitschrift für Energiewirtschaft, 32(3), 147-161.

Grimm, V., A. Ockenfels, and G. Zoettl (2008b). "Ein Vergleich ausgewählter europäischer Strombörsen" Zeitschrift für Energiewirtschaft, 32(3), 162-270.

Hart, Oliver (1995). *Firms Contracts and Financial Structure*. Oxford: Clarendon Press.

ISO New England (2004). "Final Report on Electricity Supply Conditions in New England During the January 14-16, 2004 Cold Snap." October. http://www.iso-ne.com.

Hogan, William W. and Scott M. Harvey (2000), "California Electricity Prices and Forward Market Hedging," Working Paper, Harvard University.

Joskow, Paul L. (1987). "Contract Duration and Relationship Specific Investments," *American Economic Review*, 77:168-75.

Joskow, P.L. (2005). "The Difficult Transition to Competitive Electricity Markets in the United States." *Electricity Deregulation: Where To From Here?* (J. Griffin and S. Puller, eds.), Chicago, University of Chicago Press.

Joskow, P.J. (2006). "Markets for Power in the United States: An Interim Assessment." *The Energy Journal*, 27(1), 1-36.

Joskow, P. L. (2007). "Competitive Electricity Markets and Investment in New Generating Capacity." *The New Energy Paradigm* (Dieter Helm, Editor), Oxford University Press.

Joskow, P. L. (2010). "Product Market Regulation," CESifo DICE Report, 3/2010, 3-7.

Joskow, P.L. and J. Tirole (2006), "Retail Electricity Competition," *Rand Journal of Economics*, 37(4), 799-81.

Joskow, P.L. and J. Tirole (2007). "Reliability and Competitive Electricity Markets." *Rand Journal of Economics*, 38(1), 68-84.

Kittsteiner, T. and A. Ockenfels (2006). "Market Design: A Selective Review" Zeitschrift für Betriebswirtschaft, 121-143.

Nitsche, Rainer, Axel Ockenfels, Lars-Hendrik Röller und Lars Wiethaus (2010). "Großhandelsmärkte für Strom – Marktintegration und Wettbewerb aus deutscher Perspektive." Energiewirtschaftliche Tagesfragen, 60(3): 28-37.

Müsgens, F. (2006). "Quantifying Market Power in the German Wholesale Electricity Market Using a Dynamic Multi-Regional Dispatch Model." *Journal of Industrial Economics*, 54 (4), 471.

Müsgens, F. and A. Ockenfels (2006). "Marktdesign in der Energiewirtschaft" In: Umwelt und Energie, Eds.: W. Franz, H. Hesse, H.J. Ramser und M. Stadler, Wirtschaftswissenschaftliches Seminar Ottobeuren, Tübingen: Mohr Siebeck.

Ockenfels, A. (2007a). "Strombörse und Marktmacht." Studie für das Ministerium für Wissenschaft, Wirtschaft und Verkehr des Landes Schleswig Holstein. Energiewirtschaftliche Tagesfragen, 57(5), 44-58.

Ockenfels, A. (2007b). “Marktmachtmessung im deutschen Strommarkt in Theorie und Praxis: Kritische Anmerkungen zur London Economics-Studie.” Studie im Auftrag der RWE AG. Energiewirtschaftliche Tagesfragen, 57(9), 12-29.

Ockenfels, A. (2008a). “Neun Beobachtungen zur Preisbildung im liberalisierten Strommarkt: Darf man seiner Intuition vertrauen?” In: Bonner Gespräch zum Energierecht 3, 9-29, Hrsg.: W. Löwer, Göttingen: Vandenhoeck & Ruprecht.

Ockenfels, A. (2008b). “Geht in Deutschland das Licht aus?” Frankfurter Allgemeine Zeitung. August 16, 11

Ockenfels, A. (2009a). “Marktdesign und Experimentelle Wirtschaftsforschung” Perspektiven der Wirtschaftspolitik, 10(Sonderausgabe), 31-54.

Ockenfels, A. (2009b). “Empfehlungen für das Auktionsdesign für Emissionsberechtigungen.” Zeitschrift für Energiewirtschaft, 2, 105-114.

Ockenfels, A. (2009c). “Probleme von Strompreiskontrollen aus volkswirtschaftlicher Sicht.” In Bonner Gespräche zum Energierecht, Vol. 4, ed. Wolfgang Löwer, 25-42. Bonn.

Ockenfels, Axel, Felix Müsgens and Markus Peek (2011). Observations on the Economics and Design of Power Reserve Markets. Working Paper, University of Cologne.

Oren, Shmuel S. (2005), “Generation Adequacy via Call Option Obligations: Safe Passage to the Promised Land,” *Electricity Journal*, November.

Oren, Shmuel (2005), “Generation Adequacy Via Call Options: Safe Passage to the Promised Land,” University of California Energy Institute Working Paper EPE-016, September.

Roberts, March and Michael Spence (1976), “Effluent Charges and Licenses Under Uncertainty,” *Journal of Public Economics*, 5:193-208.

Singh, Harry (2000), “Call Options for Energy: A Market-based alternative to ICAP,” October. http://stoft.com/metaPage/lib/Singh-2000-10-Options-ICAP.pdf

Sioshansi, F.P. and W. Pfaffenberger (2006), *Electricity Market Reform: An International Perspective*, Elsevier.

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

Turvey, R. (1968), *Optimal Pricing and Investment in Electricity Supply: An Essay in Applied Welfare Economics*, Cambridge, MA: MIT Press.

Vazquez, Carlos, Michel River, and Ignacio Perez Arriaga (2002), “A Market Approach to Long-Term Security of Supply,” *IEEE Transactions on Power Systems*, 17 (2): 349-357.

Williamson, Oliver (1979). “Transaction-cost Economics: The Governance of Contractual Relations,” *Journal of Law and Economics*, 22:3-61.

Wolak, Frank (2004), “What’s Wrong with Capacity Markets,” Working Paper, Stanford University. http://stoft.com/metaPage/lib/WolaK-2004-06-contract-adequacy.pdf

Wolfram, Catherine (1999). “Measuring Duopoly Power in the Deregulated UK Electricity Market,” *American Economic Review*, 89: 805-826.