

Capacity Market Fundamentals

PETER CRAMTON,^a AXEL OCKENFELS,^b and STEVEN STOFT^c

ABSTRACT

Electricity capacity markets work in tandem with electricity energy markets to ensure that investors build adequate capacity in line with consumer preferences for reliability. The need for a capacity market stems from several market failures. One particularly notorious problem of electricity markets is low demand flexibility. Most customers are unaware of the real time prices of electricity, have no reason to respond to them, or cannot respond quickly to them, leading to highly price-inelastic demand. This contributes to blackouts in times of scarcity and to the inability of the market to determine the market-clearing prices needed to attract an efficient level and mix of generation capacity. Moreover, the problems caused by this market failure can result in considerable price volatility and market power that would be insignificant if the demand-side of the market were fully functional. Capacity markets are a means to ensure resource adequacy while mitigating other problems caused by the demand side flaws. Our paper describes the basic economics behind the adequacy problem and addresses important challenges and misunderstandings in the process of actually designing capacity markets.

Keywords: Capacity markets, Market failures, Resource adequacy

<http://dx.doi.org/10.5547/2160-5890.2.2.2>

✂ 1. THE ADEQUACY PROBLEM AND WHY ENERGY MARKETS CANNOT SOLVE IT EFFICIENTLY ✂

Suppose electricity markets did not suffer from demand-side flaws. In particular, suppose demand is sufficiently responsive to prices, such that the wholesale electricity market *always* clears. Then, the market would be perfectly reliable: If supply is scarce, the price would rise until there is enough voluntary load reduction to absorb the scarcity. Consumers would never suffer involuntary rationing.¹

Yet, current electricity markets do not reflect this textbook ideal of guaranteed market clearing. The main problem is a lack of real time meters and billing and other equipment to allow consumers to see and respond to real time prices, resulting in low demand flexibility.² Because storage of electricity is costly, the supply side is also inelastic as capacity becomes scarce (capacity includes both generation and equivalent demand response, but for convenience we will often refer simply to generation). As a result, there is a possibility of non-price rationing of demand in the form of a rolling blackout, as illustrated by Figure 1. During a

1. This hypothetical case assumes automatic instantaneous demand response when needed and ignores transmission failures.

^a University of Maryland.

^b University of Cologne.

^c Corresponding author. E-mail: steven@stoft.com.

Economics of Energy & Environmental Policy, Vol. 2, No. 2. Copyright © 2013 by the IAEE. All rights reserved.

2. See Joskow (2006, 2007) and Joskow and Wolfram (2012), and the references therein, for details.

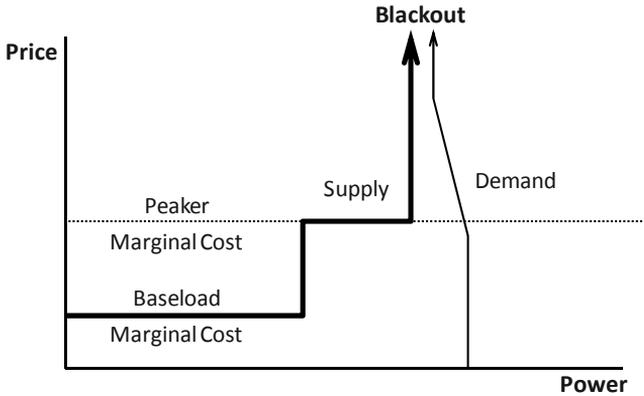


FIGURE 1

Blackouts occur when supply cannot equal demand

(rolling) blackout, all available generators produce as much electricity as they can, yet—whatever the price—not all demand can be served.

Current electricity markets do not prevent the possibility of blackouts, and the present analysis assumes that will continue to be the case.³ In fact, given the demand-side flaws, fully eliminating blackouts due to insufficient generation is unlikely to be optimal. To see this, define the “Value of Lost Load” (VoLL) as the amount that consumers would pay to avoid having supply of power interrupted during the blackout. Now suppose the average annual *Duration* of blackouts is five hours per year and that $\text{VoLL} = \$20,000/\text{MWh}$. Suppose further that the rental cost of reliable capacity (RCC) is $\$80,000/\text{MW-year}$. If one MW of capacity is added, it will run five hours per year on average and reduce the cost of blackouts by $\$100,000/\text{year}$. That is more than the cost of capacity so new capacity should be built up to the point where the duration of blackouts falls to 4 hours per year and the marginal cost of capacity equals the marginal reduction in the cost of lost load. That is, the optimal expected duration of blackouts is $\text{Duration} = \text{RCC}/\text{VoLL}$.⁴ As long as the rental cost of reliable capacity is positive, efficiency requires that blackouts occur with positive probability.

However, a key insight is that *electricity markets cannot optimize blackouts*. To see why, observe that the economics of competitive markets assumes that the price will always clear the market. That is, competitive economics starts by assuming that there is no adequacy problem (defined below), and concludes that in market equilibrium production costs are guaranteed to be minimized. However, competitive markets cannot optimize blackouts. The reason is that the duration of blackouts depends on the generation capacity built to avoid them, and the incentive to build generation to avoid blackouts depends on the price being paid during blackouts. Yet there exists no competitive market price during blackouts (Figure

3. Capacity markets generally encourage the development of demand-side resources, but even with this encouragement it appears that adequacy concerns will continue to play a significant role in electricity markets for quite some time to come.

4. We can ignore fuel costs because they are negligible compared with the rental cost of capital for a generator that runs only four or five hours a year. Also, observe that this formula does not involve the amount of unserved load, only the duration of blackouts.

1); the price that is being paid to generators during blackouts *must* be set by administrative rules.⁵

The failure of markets to optimize blackouts goes beyond the case of rolling blackouts. For instance, when capacity gets scarce there is also an increased probability of a network collapse (e.g., Joskow and Tirole, 2007; Joskow, 2008). But a network collapse implies a market collapse, because, as electricity cannot be delivered during a system collapse, consumers are not willing to pay a price during the collapse. As a result, market mechanisms cannot capture the cost of catastrophic blackouts and thus do not optimize their occurrence (Joskow and Tirole, 2007).

Observe also that the challenge to find prices during (rolling) blackouts is *not* related to the well-known literature on peak load and scarcity pricing, and investment incentives in electricity markets, starting with Boiteux (1949, 1960, 1964). Scarcity pricing relies on market clearing prices. The basic idea is that, if all available generation capacity is fully utilized, there may be 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 then required to bid prices up until the market clears. At the resulting “scarcity prices,” all generators that are supplying energy in such scarcity events earn scarcity rents, which in turn are needed to cover the fixed capital costs. This mechanism is essential to investment incentives in all energy markets (e.g., Grimm and Zoettl, forthcoming; Zoettl, 2011).⁶ But it cannot help in optimizing blackouts or in finding efficient prices when there is a possibility that no market clearing price exists due to demand-side flaws.

Peak load and scarcity pricing require high prices and electricity markets often impose “price caps.” This combination leads to the view that the root of the adequacy problem is price suppression by the regulator, and that discontinuing that price suppression can solve the adequacy problem. But this is not the case. In fact too high a price cap can result in too much capacity. The following example of how this can happen may help explain why the adequacy problem is ultimately the result of demand-side market failures and not the result of regulatory price suppression.

Suppose a blackout occurs when a large generator has been out of service for a week and the weather becomes hot and consumers gradually turn on their air conditioners. Consumers value lost load at $\text{VoLL} = \$10,000/\text{MWh}$. There is also demand elasticity with demand dropping smoothly as the price rises from $\$1,000$ to $\$20,000/\text{MWh}$, but not dropping by much.

As demand rises during the hot afternoon, it would eventually exceed total supply by a tiny amount if the price stayed at the variable cost of a peaker, assumed to be $\$200/\text{MWh}$. But instead, the price at which supply equals demand will jump to just over $\$1,000/\text{MWh}$ and that will decrease demand slightly. So far, the market is optimal and generators are earning

5. In our analyses, we assume for simplicity that generators are paid the spot price. This sometimes causes confusion because most generators sell their power forward. However, the prices for forward contracts are linked to expected spot market prices for electricity through intertemporal arbitrage. Thus, it is safe to assume power is sold only in the spot market, and we will continue to do so. Similarly, for simplicity, we mostly ignore ancillary services here, as the supply is also linked to spot market prices, and so explicitly taking them into account would not make a difference for our exposition. That said, we emphasize that the details of these markets and the protocols that guide how system operators use or do not use these markets and related options in times of scarcity, substantially contributes to the understanding of the adequacy problem. See Joskow (2007 and 2008) for insightful discussions.

6. The important role of scarcity pricing for investments in electricity markets and in addressing the adequacy problem is described in more detail by, e.g., Cramton and Ockenfels (2012). The current paper focuses on blackouts.

normal scarcity rents, as discussed above. No one is paying more than energy is worth to them. But if the regulator does not intervene, the price will continue on up to \$20,000/MWh. This is not optimal because non-elastic consumers (almost all of them) are paying twice what power is worth to them. When they overpay, as in this example, it sends a signal for the market to build too much capacity.

Again, there is no way for the market to escape this dilemma on its own. Supply and demand intersect at \$20,000/MWh and there is nothing special that any market participant can observe about the price of \$10,000 (VoLL). So, the market sets the wrong price. But as soon as demand increases another watt, supply and demand will fail to intersect at all. At that time, no price will be determined by the market. The only result that can logically be predicted is that the price might stay at its most recently determined level, \$20,000/MWh. But this is still twice too high. More importantly, the value of \$20,000/MWh set by some demand-elastic customer is not related to the average value of lost load among inelastic customers. Their average VoLL could as well have been \$1,000. The fact that just one unusual customer who is watching the price and buying wholesale has an extremely high value for giving up his last MW of power should not be relevant to determining the value of reliability for the majority of customers.

This brings us to the fundamental purpose of a capacity market, which is *to provide the amount of capacity that optimizes the duration of blackouts*. This problem is what is called the “adequacy problem.” The heart of the adequacy problem is resolving the trade-off between more capacity and more blackouts.⁷

This definition of the “adequacy problem” is convenient for at least two reasons. First, almost all observers agree that current markets do have an adequacy problem according to this definition, and they agree that this is the problem that capacity markets attempt to solve. Second, a market with an adequacy problem so defined cannot satisfy all the assumptions of perfect competition.

✎ 2. BASIC APPROACHES TO SOLVING THE ADEQUACY PROBLEM ✎

During rolling blackouts, essentially every generator is running, so all are paid the same high scarcity price. Typically, the price is capped too low. That means there is “missing money,” which implies too low a level of investment in capacity. One key observation about missing money is that, since it is missing from scarcity hours, every generator is missing essentially the same amount of money per MW of capacity. There are two basic ways to restore the missing money in proportion to MW of capacity (so that this results in incentives for building the correct mix of generation technologies): (1) raise scarcity prices paid during blackouts (price-based approach), and (2) pay every supplier of capacity the same amount per MW of

7. Alternatively, this problem could be stated as: For a market in which it is optimal to build less capacity than is required to avoid any possibility of load shedding due to lack of operable capacity, find the capacity level that optimizes the extent of load shedding. Observe that the adequacy problem is not about ‘eliminating’ blackouts and thus not about maximizing ‘security’ of electricity supply. We also note that a lack of blackouts in an energy-only market does not necessarily indicate that the market is working optimally or that blackouts have been optimized. Market power in the energy spot market will attract new entry and can even result in too much capacity, which means excess reliability. Also regulators have a number of levers with which to control investments in energy-only markets. They can (and frequently do) pay generators not to retire, pay for operating reserves, and set a high price during blackouts. The point is not that regulators should not do these things, but that if capacity is affected by such actions we are not observing a purely “competitive market” solution to the adequacy problem.

capacity (quantity-based approach).⁸ There is also a third, less commonly proposed approach which we will only briefly discuss (Section 4.b below). That is to raise the requirement of operating reserves—generators that are paid to standby and be prepared to supply more energy on short notice.

Price-based approach: energy-only market

The price-based approach uses what is often called an “energy-only” market. This is a bit of a misnomer because such markets nearly always purchase some form of operating reserve capacity, and so include capacity-based instruments. However, we will define an energy-only market to be one that attempts to solve the adequacy problem by setting a high “price cap,” which is the price paid during a blackout.

Normal market operation would dictate that the price should increase whenever demand exceeds supply. In a normal market, this will clear the market. But during a blackout, this would result in the price rising without limit since demand is very inelastic due to the basic market imperfection that causes the problem. It would also create significant opportunities for suppliers to exercise market power. Since this is not desirable, the energy price is capped. If the regulator manages to set this cap at VoLL, the market will achieve the second-best outcome, which we will, with slight exaggeration, term optimal.⁹ (This is not optimal because VoLL reflects only the *average* opportunity cost that consumers place on electricity consumption. Thus, by using this average, some consumers will be forced to buy more reliability than they want and others less, but this is the best that can be done given physical limitations.)

The market responds to VoLL by building additional capacity up to the point where a MW of capacity costs just as much as it earns from being paid VoLL during blackouts. So investment stops when the carrying cost of the last MW of capacity equals VoLL times the expected number of blackout hours, *Duration*. But $\text{VoLL} \times \text{Duration}$ is exactly the value of serving the load that would have gone unserved without that MW of generation. So at this point the cost of capacity equals the value of capacity to consumers, and beyond this point, consumer value per MW of capacity can only decline as the system becomes more reliable. Hence, the VoLL pricing rule causes the market to build the second-best, “optimal” amount of capacity. This solves the adequacy problem—with help from a regulator (Stoft, 2002; Joskow and Tirole, 2007).

The energy-only approach works because the market will build generators up to the point where an extra MW of generation makes revenues ($\text{VoLL} \times \text{Duration}$) that exactly equal its costs (*RCC*), and at that point, the equation for optimal capacity ($\text{Duration} = \text{RCC} / \text{VoLL}$) holds true.

One problem is that it is difficult to estimate VoLL (Stoft, 2002; Joskow, 2007). The reason is that current markets have hardly any access to information concerning how consumers value reliability, because consumers take few market actions that are based on reliability considerations. This is obviously true for consumers who cannot be individually interrupted, because system operators typically have no control over the electricity flows that go to individual customers. The value of reliability may be revealed only for those (large) consumers, who do have real-time meters *and* can be interrupted, *and* if system operators are prepared to

8. Because generators are not perfectly reliable, the number of scarcity hours they miss on average should be taken into account as discussed later.

9. To induce optimal capacity, the price cap must be in place not only during (rolling) blackouts but also during ‘normal’ operations (see the example for the inefficiency of prices above VoLL without blackouts at the end of Section 1).

black them out based on the performance of their suppliers (see Chao and Wilson, 1987; Joskow and Tirole, 2006, 2007). But this is of little help since it is the average VoLL of those who *cannot* respond to price that is required for the energy-only market. Thus, the price-based approach to the adequacy problem ultimately depends on the quality of the regulator's estimate of VoLL.

Quantity-based approach: capacity market

A capacity-market approach requires that the regulator calculate C^* , the level of capacity that results in the optimal duration of blackouts. This is a difficult engineering calculation, but one that regulators have historically made.¹⁰

Even with a quantity-based approach, the regulator will still need to set an energy price during blackouts, since the market cannot. However, this price will mainly serve to induce efficient behavior by existing plants and, unlike in an energy-only market, it will have no effect on the level of installed capacity. For example, assume the regulator sets a low price of, perhaps, $P_{CAP} = \$1,000/\text{MWh}$. This will be too low to induce an optimal capacity level of C^* .

To illustrate the fundamental difference of the quantity-based approach, first consider the capacity-market design that is most similar to the price-based approach. In this design the capacity market is used to top-up the energy price to the level that induces C^* . The regulator holds an auction for C^* MW of capacity and allows new and existing capacity to bid a scarcity price, P_s , (a price during blackouts) that would induce generators to remain in or to enter the market. The lowest price, P_s^* , that would be accepted by at least C^* of capacity would become the market's new price cap. Then during scarcity hours, the capacity market would pay all generators that sold capacity, $P_s^* - P_{CAP}$, on top of the energy-market payment of P_{CAP} . In effect, the auction discovers the value of the price cap that would correspond to C^* and implements that as the energy price during blackouts.

This solves the adequacy problem, and it avoids market coordination problems that occur when the market builds capacity in response to energy prices instead of a capacity auction. The auction coordinates the investors decisions to build, so they neither under- nor over-build. While we will suggest a different design (see next section), this demonstrates that a capacity market can act just like an energy-only market except for giving the regulator control over capacity.¹¹

The reverse approach to designing a capacity market is equally simple. Instead of capacity suppliers bidding for a higher scarcity price, P_s , they could bid for a capacity payment, C_{PAY} . As we will discuss in the next section, the advantage of this approach is that it does not increase risk and market power the way increasing the peak energy prices does.

For the regulator, the price and quantity approaches differ, because the regulator determines the value of lost load (VoLL) in one and determines C^* in the other. Since these parameters control the capacity level and the duration of blackout, these two approaches are equally regulatory in nature.

10. In reality both the determination of VoLL and the optimal duration of blackouts will likely remain a highly politicized process. There is no objective way to estimate VoLL accurately, and whoever will be held responsible when blackouts occur will want and usually obtain influence over the selected values.

11. In particular, note that this design provides exactly the same efficient real-time signals to build generators for both adequacy and security blackouts as does an energy-only market. (A security blackout is one that is triggered by an unanticipated line or generation outage and that could have been avoided if the system operator had dispatched more existing generation in advance.) So there is no reason for the frequent claim that capacity markets should avoid helping with security problems.

With the quantity approach, C^* can be determined either from a target duration of blackouts, or it can be derived from VoLL. But even if derived from VoLL, the first step is to estimate a duration of blackouts from $Duration = RCC/VoLL$. Then, since blackout probabilities depend on the amount of installed capacity, C , it is possible to back out the value C^* that is the level of C that causes the desired duration of blackouts. The calculation of blackout probabilities is a difficult task, but one that engineers have decades of experience with since regulated utilities use essentially the same approach to decide how much capacity to build.

Traditionally, and often with capacity markets, a target duration such as “one day in ten years” is used. “One day” is sometimes taken to mean 24 hours of blackouts and sometimes taken to mean one event of, perhaps, three hours. That discrepancy gives an indication of the arbitrariness of the target. However, differences in the cost of electricity under those two standards are actually quite small, perhaps less than one percent, because spare peaking capacity is relatively cheap to build or keep online and because it requires essentially no fuel and few additional power lines.

In summary, the choice between the two basic approaches, price and quantity, is not a choice between a market approach and a regulated approach. And both the quantity and price approach can solve the adequacy problem. So the choice between the two depends on other factors, such as risk, market power, and the coordination of investments in capacity.

✎ 3. A PRACTICAL APPROACH TO SOLVING THE ADEQUACY PROBLEM: ✎ RELIABILITY OPTIONS

This section describes design features of a capacity market, based on experience in actually designing capacity markets such as in Colombia, New England and as being considered in UK and Germany, as well as on converging recommendations for capacity market designs as surveyed by Cramton and Stoft (2006).¹²

The advantage of capacity payments

The adequacy problem can be solved either by setting the price cap to VoLL or by adding a capacity market that targets C^* . Assuming the market has an adequacy problem, there are several reasons to select capacity mechanisms.

An obvious practical reason to use a capacity market is that circumstances do not permit a price cap to be credibly set at VoLL. It may be politically difficult to allow the price of a MWh that normally sells for \$40 to reach \$20,000 simply because some committee has estimated that \$20,000/MWh is the average value of lost load. And even if this is allowed initially, investors may not believe that the policy is durable, in which case it will not induce the required investment. If a high price cap is not feasible, a capacity market is the preferred choice. But, as we will see, there are also serious risk and market power issues with high price caps. A more refined approach, one that addresses such risk and market power issues, adds reliability options to the market.

12. Related approaches are described in Bidwell (2005), Chao and Wilson (2004), Cramton and Stoft (2006, 2007, 2008), Oren (2005), Vazquez et al. (2002), and Cramton and Ockenfels (2012).

Reliability options

To give a brief overview, the capacity market we describe in more detail below 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 prices capacity from the bids of competitive new entry in an auction. Two major advantages of reliability options are that the capacity payment (a) hedges load from high spot prices and (b) reduces supplier risk by replacing peak energy rents (the rents derived from selling energy at 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 (see, e.g., Cramton and Stoft (2006) for expanded discussions).

To be more specific, the reliability options are introduced into the market by requiring every generator that receives a payment for X MW of capacity to sell a reliability option for X MW of capacity. The specific form of reliability options that has been implemented in practice is *load following* reliability options. This reduces option obligations in proportion to reductions in load, and thereby minimizes risk for both generators and load by preventing over-hedging of load, so it is the one we recommend and will discuss.

The options will have a strike price of, perhaps, \$300/MWh. In this case, whenever the spot price, P , is above \$300/MWh the generator must pay load $(P - \$300)/\text{MWh}$. From a financial point of view the price the load and generators face is capped at \$300/MWh. However, reliability options leave incentives in a competitive market fully intact: 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 (load following) obligation, it is paid \$80,000 because the spot price is \$1,000, but it must pay $(90 \text{ MW}) \times (\$1000 - \$300)$ 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 is profitable 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. And load, too, is 100% hedged from energy prices in excess of the strike price.

The hedging mechanism can also be explained under the simplifying assumption that the energy price stays low except when there is a blackout due to a shortage of capacity (an adequacy problem), at which time it rises to the price cap, P_{CAP} . Load may be at various levels when such a blackout occurs because the amount of generation out of service varies. So a reliability option assigns generators a capacity “load share,” C_{LS} , which is the same fraction of total load as their capacity is of total capacity.

$$C_{\text{LS}} = C_{\text{BID}} \times (\text{Load Served}) / \sum C_{\text{BID}} \quad (1)$$

The first C_{BID} is the accepted bid of the generator with load share equal to C_{LS} . We will assume that all generators in the energy market have had their bids accepted and so $\sum C_{\text{BID}}$ is the sum of the designated capacity value of all generators in market. This means that

$$\Sigma C_{LS} = \text{Load Served.} \quad (2)$$

The financial option associated with a reliability option specifies that during a blackout, a generator must pay C_{LS} times $(P_{CAP} - P_{STRIKE})$ to load, but the energy market pays the generator its output, C times P_{CAP} . It is convenient to compare such a capacity market to an energy-only market capped at P_{STRIKE} , which pays $C \times P_{STRIKE}$. In this case, the energy payments of a generator with a reliability option minus the energy payments in the capped energy-only market are

$$\text{Net Energy Payments with a Reliability Option} = (P_{CAP} - P_{STRIKE})(C - C_{LS}) \quad (3)$$

If we sum these over all generators, the last term becomes $(\text{Total Output} - \text{Load Served})$, which is exactly zero because of physics as noted by equation (2). Consequently, relative to a market capped at P_{STRIKE} , energy payments in such a capacity market are exactly zero. This means load and total generation are both perfectly hedged relative to a market with a low price cap.

As will be discussed later, reliability options not only reduce risk but also market power without damaging the dispatch incentives on the generation side. However, a capacity auction is needed to set the payments to generators for providing reliability options just high enough to induce adequate capacity.

These auctions determine the price of reliability options that is just sufficient to induce the required entry.¹³ 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 optimal spot market without reliability options. 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 fully 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 is limited to the determination of the one factor the market cannot control—the adequate level of capacity.

Missing money and the size of the capacity payment

The reliability option approach also needs to specify a price cap, P_{CAP} , on the spot market for times of energy shortage, which should reflect the best estimate of VoLL. This will, however, play little role in determining investment, and so the exact level is not critical.

To be more specific, observe that if a generator produced 1 MW more during all shortage hours, it will then expect to earn more by an amount called the *peak energy rent* (PER), which is the difference between the price cap and the strike price, $P_{CAP} - P_{STRIKE}$, times the expected number of shortage hours in a year. The ideal PER, denoted by PER^* , is the same except with P_{CAP} set to VoLL. The difference between the two is the missing money of a perfectly reliable megawatt of capacity, so if we multiply that by the average load served during a blackout, L_{AS} , we find the total annual missing money, MM, measured in \$/year:

13. All capacity markets have in common that there are capacity auctions that select the capacity that is willing to enter or remain in the market for the lowest possible capacity payment. The auction must prevent the exercise of market power. Although auction design is a crucial and controversial part of capacity-market design, it is less fundamental than the considerations this paper addresses, so we will not discuss its design or role further (but see Cramton and Stoft 2006 and Cramton and Ockenfels 2012).

$$MM = (PER^* - PER) \times L_{AS}, \quad (4)$$

In order for the capacity market to induce the optimal amount of capacity, it must replace the missing money. It does this by paying generators for their capacity, C_{BID} , times the capacity price, C_{PRICE} . That product, the total capacity payment to a generator, will be called C_{PAY} . The total of all such payments is more than the missing money because capacity payment must cover both missing money and the cost of the ROs that the generators sell. But that cost is just $PER \times L_{AS}$, so another formula for the missing money is

$$MM = \sum C_{PAY} - PER \times L_{AS} \quad (5)$$

Equating the right sides of equations (4) and (5) gives the formula for total annual capacity payments:

$$\sum C_{PAY} = PER^* \times L_{AS} \quad (6)$$

This tells us that total capacity payments do not depend on missing money, MM, but only on the ideal peak energy rent and the average load during blackouts. So as the missing money problem is reduced, the capacity payments will remain constant¹⁴ because as the price cap is raised the cost of reliability options goes up as fast as the amount of missing money decreases. However, note that as the reliability option becomes more costly it saves load as much money as it costs, and it becomes a more valuable policy by preventing the increased risk and market power that would come with higher scarcity prices.

Setting the capacity rating, C_{BID}

So far, we have not addressed the question of what determines C_{BID} . It would be desirable for investors to be motivated to bid the quantity of capacity honestly because, for old plants and demand-side resources, they will have a more accurate view of its value than does the system operator. Also, investors will lobby to be allowed to determine C_{BID} . However, as long as there is a significant amount of missing money, investors will want to set C_{BID} as high as possible.

Setting C_{BID} is a particularly difficult problem for demand-side resources for two reasons. First, their contribution to reliability is to not use electricity and “not-using” must be measured in comparison to the hypothetical value of “what they would have used had they not been contributing to reliability.” This measurement requires the use of statistics and a baseline that is not easy to game. Second, “not using electricity” is already rewarded by not paying for the unused electricity, and this must be properly accounted for to avoid over payment. But as long as these problems are resolved adequately, demand-side resources can be some of the most inexpensive for supplying capacity.

To see why investors will set C_{BID} as high as possible, consider a “cardboard plant” that produces nothing, and which has sold C_{BID} of capacity. How much can it expect to make? When it sells capacity its payment is in part for the reliability option and in part to compensate for missing money. Because it produces nothing, it should expect to lose the part of the payment for the reliability option. But it will still be left with its share of the missing money:

$$\text{Net Payments to a Cardboard Plant} = MM \times C_{BID} / \sum C_{BID}. \quad (7)$$

14. Of course, this is just the theoretical expected value of capacity payments, and actual payments will depend on whether the auction buys new capacity, on variations in expectations of future missing money, and various other factors.

So as long as there is significant missing money, increasing C_{BID} will be expected to increase a generator's net payment from the reliability option it has sold. For example, if P_{CAP} is only a little above the strike price, there will be a lot of missing money and there will be little cost to selling a financial option. In this case, claiming that your plant has 100 MW of reliable capacity when it actually has only 80 MW will be just as profitable as telling the truth about this plant but claiming you have another 20 MW plant when it is only a cardboard plant. And as equation (4) shows, that will be quite profitable. This is why the system operator needs to set C_{BID} and why it may be wise to adjust a unit's C_{BID} rating based on historical performance in scarcity events, as is done in New England.

Typically, investors will want the highest possible C_{BID} , but what is the right C_{BID} to assign them? This value should not inappropriately favor one type of generator over another; otherwise the market will select the wrong mix of capacity types. So C_{BID} should be set so that the capacity market mimics a perfect energy-only market. This means that every MW of capacity should be paid the same on average as they would be paid by optimal scarcity prices, taking account of the fact that some generators are more reliable than others. If a generator is available twice as often during scarcity events, then it would receive twice as much peak revenue from an energy-only market. So it should also receive twice as much from a capacity market.

So, in order to properly reward capacity for its contribution to adequacy, it is necessary to set C_{BID} to reflect the actual contribution of generating units to adequacy. (In a more complete model, we would find that C_{BID} should reflect a generator's contribution to all scarcity events including those caused by security issues.) This means setting C_{BID} equal to nameplate capacity times the probability that the unit will provide capacity during a scarcity event.

Another reason for the system operator to determine the capacity ratings to be used in the auction is to achieve consistency between the estimate of needed capacity and the amount of capacity purchased. For example, the system operator might find that a reliable market requires the 100 existing generators, with an estimated effective capacity of 30 GW, plus another gigawatt of new capacity. If generators determined their own C_{BID} values, they would likely inflate them and when the auction bought 31 GW of capacity it would find this consisted of 90 of the existing generators and no new ones. At that point, 10 existing generators might retire and the system would become unreliable. Using the same capacity values for determining required capacity and for the purchase of capacity eliminates such discrepancies.

Energy market power and risk

Having covered the mechanics of the reliability option, we now consider its benefits. As noted above, its actual purpose is neither to provide performance incentives nor to procure capacity, but rather to help a basic capacity market improve on the performance of the high-price-cap energy markets by reducing market power and risk.

First, consider market power. Suppose the strike price is \$350, and the marginal cost of generation is \$50. Suppose that an investor with 1 GW of generation can withhold 100 MW and drive the price up to \$3,050. In this case, its profit increases from $1000 \text{ MW} \times \$300$ to $900 \text{ MW} \times \$3000$, an increase in profit of \$2,400,000 per hour. But with a reliability option, if its load-share capacity is 1 GW (1000 MW), then when it reduces its output to 900 MW, it will owe \$2,700 on the 100 MW it is short of its hedging requirement of C_{LS} , for a net loss of \$270,000 per hour. Clearly a huge incentive to exercise market power has been elim-

inated by the reliability option and turned into a disincentive. This is a standard result of selling power forward, which is what a reliability option does (e.g., Allaz and Vila, 1993; Hogan and Harvey, 2000; Cramton and Stoft, 2008). Basically, a reliability option insures that generation will sell its on-peak energy forward at a competitive price determined in the capacity auction thereby eliminating its excessive market power during scarcity events.

We now turn to the question of market price risk (price volatility). A completely reliable generator will experience a great deal of risk in an energy-only market because the number of scarcity hours with the price at the price cap will vary dramatically from year to year (Stoft, 2002; Cramton et al., 2006; Joskow, 2007; Cramton and Ockenfels, 2012). However, with a reliability option, scarcity payments from the energy market will be almost entirely hedged away and replaced with capacity payments that do not vary with the number of annual scarcity hours. Hence a reliability option will greatly reduce market risk. It will not, however, reduce unpredictable performance risk. Nevertheless, the investor's risk premium in a capacity market should be significantly lower than in an energy-only market, and this should save consumers money.

The primary misconception reliability options engender is the belief that individual generators and loads feel as if they are price capped. With respect to performance incentives, this represents a failure to understand options. It is only generation as a whole that is capped, not individual generators. Their incentives are undiminished, because they still face the full spot-market price on the margin, even though their risk is hedged. Yet, while leaving performance incentives untouched the financial option manages to reduce market power exactly *as if* there were a price cap at P_{STRIKE} . Such a low cap would indeed be highly effective at reducing market power, and a reliability option is just that effective.

✎ 4. ALTERNATIVE APPROACHES TO ADDRESS THE ADEQUACY PROBLEM ✎

Demand elasticity

There are more ways to address the adequacy problem than the ones we discussed. The adequacy problem could be eliminated by increasing *demand elasticity* to the point where the energy price never exceeds the value of energy to the average customer. In such a world, load could fully protect itself against blackouts, and could somewhat mitigate market power in times of scarcity through increased demand response. While circumstances could change, for instance, with the prevalence of smart grids, smart metering, and real-time pricing, as long as demand remains rather inflexible it cannot fully mitigate adequacy problems at scarcity events. Moreover, increased demand responsiveness may not fully eliminate missing money, risk and market power problems associated with scarcity events. So, even as the market moves towards an ideal environment, a capacity market could still be beneficial.

Operating reserves

Another approach uses a low price cap and solves the adequacy problem by buying *operating reserves*. This has not been much discussed in the literature, but we note that capacity bids in 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 investment costs of new capacity,

thus cannot efficiently incentivize new entry, and so cannot 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 market, increasing the wholesale price and thus creating incentives for new investment. However, it seems unlikely that creating such incentives through stricter operating reserve requirements is an efficient way of solving the adequacy problem. Holding back capacity from the wholesale market leads to an inefficient dispatch, and can bias the long-term technology-mix. Also, it seems difficult to compute reserve requirements that would eventually incentivize the efficient amount of total capacity.

Strategic reserves

Another option which has been often put forward to address the adequacy problem is sometimes called a *strategic reserve*. This can take many forms. In the simplest version, regulators subsidize generators in order to induce their construction or to keep them in the market. One obvious disadvantage is that the strategic reserve often lets regulators choose the mix of technologies or the location of the strategic reserve, while both capacity markets and energy-only markets allow the market to decide.

More sophisticated versions, as e.g. implemented in Sweden and currently considered in Germany, procure generation capacity and let it bid into the wholesale market only if otherwise market clearing is not possible. This is a hybrid between price-based and quantity-based approaches. Yet it is inefficient. First, if the strategic reserve is not allowed to submit bids in normal times, the dispatch will be inefficient. Second, if the bids of the strategic reserves are large, probably approximating VoLL, we run into similar problems as those discussed for price-based approaches. Moreover, there is no model that we are aware of that predicts how the non-strategic capacity changes as a function of the strategic reserve. For example, if the strategic reserve is dispatched at a price determined by the last commercial bid only (as is the case in Sweden), this removes all scarcity rents and thus even suppresses new investments and distorts prices away from efficient spot market pricing. The likely result is that total capacity is reduced by the strategic reserve.

✎ 5. FURTHER PRACTICAL ISSUES AND COMMON MISUNDERSTANDINGS IN ✎ PRACTICAL DESIGN

Do renewables add to the problem?

Does the adequacy problem get more severe with an increasing share of renewables? Since renewables do not change the value of lost load (though they may affect the amount of lost load), they cannot increase the appropriate level of the price cap, so renewables do not affect our first reason for capacity markets—too low a price cap.

However, renewables may contribute to the adequacy problem. In Germany, for instance, renewables are subsidized via a feed-in tariff which is fixed and guaranteed for 20 years plus an obligation of system operators to feed this electricity into the grid whenever it is produced. This partly explains why renewables' supply is not price-sensitive (another reason is that incremental costs are low). As a consequence, renewables create the same problem as the problem which is at the heart of the adequacy problem: price-inelastic demand. In fact, renewables can be thought of as completely price-inelastic *negative* demand.

Moreover, because neither wind nor the sun can provide firm energy, renewables can only, in small part, substitute for 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.¹⁵ This makes investments in conventional resources, *ceteris paribus*, less attractive, in particular when renewables are planned to produce a considerable share of consumed electricity. 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. As a result, investors face a large amount of uncertainty with regard to future energy prices, the needed future mix of generation, and future regulatory interventions. The capacity-market benefit of market coordination may then become useful.

Why pay existing plants?

There is confusion around giving capacity payments to generators that would remain in the market without such payments. Why are they being paid to do what they would do anyway? The simplest answer is one that answers a good many, perhaps even most, capacity-market puzzles. Existing plants are paid because that is what an ideal energy-only market would do. An energy-only market pays for peak capacity with scarcity prices that rise above the variable cost of a peaker. But when there is scarcity, every operable plant is running, and all are paid the same scarcity price—even though they would stay in the market without such payments.

A deeper answer would explain why ideal markets pay existing plants the same for their electricity as new plants. One could imagine inducing new plants to enter the market with long-term contracts that pay for both variable and fixed costs, but paying existing plants only a tiny bit more than their variable costs. They would not close, because some profit is better than none. This is called a regulatory taking or expropriation, and it can work effectively if investors are surprised. But once the policy is known, new plants will demand contracts that protect them from future expropriations and likely charge a significant risk premium as well.¹⁶

Won't capacity payments distort the technology mix?

A deep confusion concerning capacity markets is that they will not induce the right type of capacity. This confusion takes many forms and contains contradictory views, but there is one basic explanation of why the standard design induces the right technology mix. It does what an ideal energy-only market does.

Start with an energy only market that sets a price cap of VoLL. Peakers will cover fixed costs only when the price exceeds their variable cost. But baseload plants also cover fixed costs when the spot price equals the variable cost of peakers but exceeds their own variable costs. This yields the well-known result that the market induces the optimal mix of plant with differing fixed and variable costs.

Now suppose we choose to reduce the price cap to \$1,000/MWh. This will result in “missing money,” and that money will be missing from scarcity hours and only from scarcity

15. Which is what is currently observed in Germany, where even very efficient gas-fueled generators struggle ‘to be in the money’ (and are now partly subsidized to not leave the market). At the same time, EWI (2012) comes to the conclusion that a substantial amount of new gas turbines needs to be built in the coming years to back up the Energiewende, although it is predicted that many of these turbines are not called at all in a ‘typical’ year.

16. Some argue that strategic reserves are preferable because capacity payments are only paid to the strategic reserves. However, if the strategic reserves submit high bids in times of scarcity, all existing plants will benefit, too—just like with all other proposals.

hours. Suppose that a perfectly reliable generator would be missing \$50,000/MW. In this case an 80% reliable generator would be missing \$40,000/MW. Now let us implement a capacity market to replace this missing money. This market will not replace the money during scarcity hours, but will instead dole it out in equal monthly payments. Won't this distort the market's choice of technologies?

In the capacity market, a perfectly reliable 1 MW generator will be paid \$50,000 per year, and an 80% reliable 1 MW generator will be paid for 0.8 MW of capacity and so it will receive \$40,000 per year. In short all of the missing money will be restored to the same generators that lost it when the price cap was lowered. Hence the capacity market will have no effect on the technology mix relative to the ideal energy-only market. This result can be seen more easily if we assume all generators are perfectly reliable. In this case, all are running during scarcity events and all lose the same amount of missing money and all receive identical capacity payments.

Adequacy versus security

The motivation for a capacity market always stems from a concern for adequacy—having sufficient operable capacity to meet demand almost always. This goal can be achieved just as surely by purchasing only gas turbines or by purchasing an efficient mix of capacity. Having an efficient mix of capacity, as opposed to too many gas turbines, does not address the adequacy problem in any way. But it seems foolish to buy capacity that does not minimize consumer costs simply because efficiency was not a motivation for the capacity market.

This reasoning—that we should strive for efficiency—is considered uncontroversial until it is noted that the mix of, and behavior of, capacity should address security concerns. Then a chorus arises to assert that security was not the motivating problem so the capacity market should ignore it.¹⁷ But providing security is just a matter of minimizing the total cost to consumers of generation and blackouts. Why should this particular cost minimization be ignored? There is no logic to this view, so “reasons” are invented.

The two most popular “reasons” to ignore security concerns are that (1) addressing such concerns will cause most generators to stay warm all the time, and (2) security is already handled optimally by the system operator. To prevent the capacity market from addressing security it is suggested that the reliability option only apply when a warning is issued, say, four hours in advance of the scarcity price. This way generators that fail to perform without a four hour warning will not owe any option payment.

But let us look carefully at what happens with a four-hour delay. First consider a standard reliability-option market. With or without the delay, if a generator supplies power in a scarcity hour it will be paid the scarcity price. The chance of this is what motivates generators to stay warm, so that motivation is not changed by a four-hour warning requirement. What could be changed is the hedge payment. But the hedge payment provides no incentive to perform because that payment (from generators to load) must be made whether or not a generator performs. If there is a scarcity hour, the generator must pay $C_{LS} \times (P_{CAP} - P_{STRIKE})$, no matter what. This cannot cause a change in behavior except through psychological thinking that seems unlikely: “If I must make a hedge payment, then I am keen to make some money

17. These arguments have played prominent roles in the US and they triumphed in the first-round UK design. However a compromise was then adopted that allows the capacity market to reward a few generators and storage devices that ramp up quickly during an unexpected event.

selling expensive power to compensate for that distress. But with no hedge payment, I will pass up this lucrative opportunity.”

What is really going on with this argument is that slow generators know they may lose out on energy market payments. As always, they dislike this possibility and would like someone to subsidize their loss. Since the reliability option requires them to make option payments, cancelling those payments is a way to disguise the subsidy—an opportunity for obfuscation not provided by an energy-only market. They then argue that if they are not given the subsidy, bad things will happen in the energy market, even though cancelling their hedge obligation has no impact on their performance.

Now consider the market design currently proposed in the UK. The two “reasons” have more surface plausibility in this case. In this design the capacity market augments the exceptionally low prices in the energy market, thus making it more like an energy-only market. This increases performance incentives by making the market more like an optimal energy-only market. And a four-hour-warning requirement would turn off these energy-market prices that bring that market closer to the optimal design. So the two “reasons” for the warning are, in essence, claims that optimal energy-only markets are detrimental because they cause generators to over-perform with regard to system security. For some reason, this view is never expressed with regard to energy-only markets, but is widely held whenever the discussion of capacity markets turns to real-time performance.

Trading between different zones

In an interconnected electricity market the actions of any element in the system impacts all others.¹⁸ 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, 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 bids into a German capacity auction. Observe that, according to the definition of reliability options, 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

18. This subsection is taken from Cramton and Ockenfels (2012).

in one market without reliability in the other market. Thus, in this specific case of fully integrated markets, it is important to align the design and implementation of a capacity market. That said, we emphasize that a well-designed capacity market is an efficient way of guaranteeing resource adequacy, 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 with an obligation of Q_{Share}) 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, $Q_{\text{Max}} - Q_{\text{Domestic}}$, and the settlement works as follows.

$$\begin{aligned} \text{Generator Revenue} = & P_{\text{Forward}} \times Q_{\text{Forward}} + P_{\text{Strike}} \times (Q_{\text{Share}} - Q_{\text{Forward}}) \\ & + P_{\text{Balance}} \times (Q_{\text{Domestic}} - Q_{\text{Share}}) + P_{\text{Export}} \times (Q_{\text{Max}} - Q_{\text{Domestic}}) \end{aligned}$$

Because the derivative of Generator Revenue with respect to Q_{Domestic} is $(P_{\text{Balance}} - P_{\text{Export}})$, 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 $\text{Max}(0, (P_{\text{Balance}} - P_{\text{Strike}}))$ 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.

$$\text{Peak-Load Penalty} = \text{peak energy cost} - L_{\text{Share}} \times (\text{Sum of all peak energy costs})$$

The sum of the penalties is zero, and the derivative of a load's penalty with respect to its own peak energy cost is $(1 - L_{\text{Share}})$, 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 $P_{\text{Strike}} + (1 - L_{\text{Share}}) \times (P_{\text{Balance}} - P_{\text{Strike}})$, which is 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. They also know 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.

✎ 6. CONCLUSION ✎

There is an increasing concern by policy makers in Europe and elsewhere that liberalized wholesale markets for electricity do not provide the incentives to build adequate generation capacity. The concern is justified. Current electricity markets with their demand-side flaws cannot determine efficient prices that would minimize total costs, including blackout costs. The central problem is the failure of wholesale markets to generate prices that reflect the opportunity cost that consumers place on electricity consumption in times when all available capacity is fully utilized. Yet these prices are crucial for the decision to invest, or not, in generation capacity.

This paper documents the fundamentals of a forward capacity market with a reliability option, which can address the adequacy problem. The fundamental points are that a well-designed capacity market has two key similarities with a well-designed energy-only market, and two key differences.

The similarities are that both designs rely on regulators or the government to determine the level of reliability that the market will provide, and both designs utilize the same high spot prices to provide performance incentives.

The differences come from the capacity auction and the financial option. The auction coordinates investment and reduces the risk of over- and under-investment, risks that are ultimately paid for by customers. Reliability options have the same effect on market power as a low price cap set at the option's strike price, but without the incentive distortions of a price cap. Reliability options also reduce energy-market risk, just as a price cap would, but they do not reduce performance risk, because they do not reduce performance incentives. In contrast, an energy-only market that is strong enough to solve the adequacy problem suffers from dangerous levels of market power and imposes significant costs of risk on consumers. Thus, the market power and risk reduction provided by a capacity market is apt to be significant.

However, the advantages come at a cost. A capacity market is an ideal target for industry lobbying because it can, and sometimes has been, turned into a system of low-risk subsidies

for existing generation. This vulnerability results from market complexities (though it is less complex than a real-time spot market) and from a misunderstanding of the fundamentals of capacity market design.

✦ ACKNOWLEDGMENTS ✦

This paper builds on earlier work, which often provides much expanded discussions and derivations, including Cramton and Stoft (2005, 2006, 2008, 2010), Joskow (2007, 2008), Joskow and Tirole (2006, 2007), Bidwell (2005), Stoft (2002), Ockenfels (2007a, 2008a, 2008b), Cramton and Ockenfels (2012), as well as on our various practical experiences in designing capacity, reserve and other energy markets. Ockenfels thanks the German Science Foundation for support through the Leibniz program and the research unit “Design & Behavior” (FOR 1371).

References

- Allaz, B. and J.-L. Vila (1993). “Cournot competition, forward markets and efficiency.” *Journal of Economic Theory*, 59(1): 1–16. <http://dx.doi.org/10.1006/jeth.1993.1001>.
- Bidwell, M. (2005). “Reliability Options. A market-oriented approach to long-term adequacy.” *Electricity Journal*, 18(5): 11–25. <http://dx.doi.org/10.1016/j.tej.2005.03.010>.
- Boiteux, M. (1949). “La tarification des demandes en pointe: application de la théorie de la vente au coût marginal.” *Revue générale de l'électricité*, 58: 321–340.
- Boiteux, M. (1960). “Peak Load Pricing.” *Journal of Business*, 33: 157–79 [translated from the original in French published in 1949]. <http://dx.doi.org/10.1086/294331>.
- 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.
- Chao, H.P. and R. Wilson (1987). “Priority Service: Pricing, Investment and Market Organization.” *American Economic Review*, 77(5): 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. <http://dx.doi.org/10.1023/A:1020535511537>.
- Chao, H.P. and R. Wilson (2004). “Resource Adequacy and Market Power Mitigation via Option Contracts.” EPRI, Palo Alto, CA.
- Cramton, P. and A. Ockenfels (2012). “Economics and Design of Capacity Markets for the Power Sector.” *Zeitschrift für Energiewirtschaft*, 36(2): 113–134. <http://dx.doi.org/10.1007/s12398-012-0084-2>.
- Cramton, P. and S. Stoft (2005). “A Capacity Market that Makes Sense.” *Electricity Journal*, 18: 43–54. <http://dx.doi.org/10.1016/j.tej.2005.07.003>.
- Cramton, P. and S. Stoft (2006). “The Convergence of Market Designs for Adequate Generating Capacity.” White paper for the California Electricity Oversight Board, March 2006.
- Cramton, P., S. Stoft, and J. West (2006). “Simulation of the Colombian Firm Energy Market.” Working Paper, University of Maryland.
- Cramton, P. and S. Stoft (2007). “Colombia Firm Energy Market.” *Proceedings of the Hawaii International Conference on System Sciences*.
- Cramton, P. and S. Stoft (2008). “Forward Reliability Markets: Less Risk, Less Market Power, More Efficiency.” *Utilities Policy*, 16(3): 194–201. <http://dx.doi.org/10.1016/j.jup.2008.01.007>.
- EWI (2012). “Untersuchungen zu einem zukunftsfähigen Strommarktdesign.” Report for BMWi.
- Grimm, V. and G. Zoetl (forthcoming). “Investment Incentives and Electricity Spot Market Competition.” *Journal of Economics and Management Strategy*.
- Hogan, W. and S.M. Harvey (2000). “California Electricity Prices and Forward Market Hedging.” Working Paper, Harvard University.
- Joskow, P.J. (2006). “Markets for Power in the United States: An Interim Assessment.” *The Energy Journal*, 27(1): 1–36. <http://dx.doi.org/10.5547/ISSN0195-6574-EJ-Vol27-No1-2>.

- 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. and J. Tirole (2006). "Retail Electricity Competition," *Rand Journal of Economics*, 37(4): 799–815. <http://dx.doi.org/10.1111/j.1756-2171.2006.tb00058.x>.
- Joskow, P.L. and J. Tirole (2007). "Reliability and Competitive Electricity Markets." *Rand Journal of Economics*, 38(1): 68–84. <http://dx.doi.org/10.1111/j.1756-2171.2007.tb00044.x>.
- Joskow, P.L. (2008). "Capacity Payments in Imperfect Electricity Markets: Need and Design." *Utilities Policy* 16(3): 159–170. <http://dx.doi.org/10.1016/j.jup.2007.10.003>.
- Joskow, P.L. and C. Wolfram (2012). "Dynamic Pricing of Electricity," *American Economic Review Papers and Proceedings*, 102(3): 381–385. <http://dx.doi.org/10.1257/aer.102.3.381>.
- Ockenfels, A. (2007a). "Strombörse und Marktmacht." *Energiewirtschaftliche Tagesfragen*, 57(5): 44–58.
- Ockenfels, A. (2007b). "Marktmachtmessung im deutschen Strommarkt in Theorie und Praxis: Kritische Anmerkungen zur London Economics-Studie." *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., F. Müsgens, and M. Peek (2011). Observations on the Economics and Design of Power Reserve Markets. Working Paper, University of Cologne.
- Oren, S.S. (2005). "Generation Adequacy via Call Option Obligations: Safe Passage to the Promised Land." *Electricity Journal*, 18(9): 28–42. <http://dx.doi.org/10.1016/j.tej.2005.10.003>.
- Stoft, S. (2002). *Power System Economics*, IEEE Press. <http://dx.doi.org/10.1109/9780470545584>.
- Vazquez, C., M. River, and I. Pérez-Arriaga (2002). "A Market Approach to Long-Term Security of Supply." *IEEE Transactions on Power Systems*, 17(2): 349–357. <http://dx.doi.org/10.1109/TPWRS.2002.1007903>.
- Zoettl, G. (2011). "On Optimal Scarcity Prices." *International Journal of Industrial Organization*, 29(5): 589–605. <http://dx.doi.org/10.1016/j.ijindorg.2011.01.002>.