Consumer Welfare Implications of Capacity Markets in Liberalized Electricity Sectors

Though it has been shown that consumers pay higher electricity prices in areas with capacity markets, those markets also serve as an insurance mechanism in incentivizing capacity additions and thereby reducing the probability of extreme events. Critics argue that the lack of a capacity mechanism in Texas ERCOT has resulted in a dangerously shrinking capacity reserve margin.

I. Introduction: A Tale of an Island

Once upon a time there was an island inhabited by 100 happy citizens. To travel from the island to mainland, the islanders needed to cross a bridge. The bridge was quite narrow, and only one person could use it at once. It usually took one minute for a person to cross it. Furthermore, there was a toll or “per-travel” fee to cross the bridge equal to $1, so each time a person crossed it she needed to pay $1. The fee was used to cover the maintenance costs of the bridge.

In a celebrated speech, the president of the island proposed to build a newer and wider bridge, so more than one person would be able to travel to the mainland at a time. The president proposed that raising the current toll could fund the new bridge. However, the citizens of the island disliked this idea. Indeed, they were smoothly coordinated in crossing the bridge, so it was rare for them to
wait more than a minute to use it. Therefore, they were not willing to pay more than $1 to cross it. A few days after the president’s speech, the islanders received a catastrophe alert. The island had to be evacuated in less than one hour in order to avoid the horrible consequences of the catastrophe. At the moment of receiving the alert, all the islanders went to the bridge to cross it immediately. However, they were aware that not all of them would be able to cross the bridge before the catastrophe hit the island. Desperate, some islanders offered to pay their peers huge amounts of money to cross the bridge before them. Other islanders asked themselves, “Why didn’t we use this money that some of us are willing to pay to cross the bridge now to build the new bridge that the president suggested to us?”

II. Capacity Markets at a Glance

The president of the island was anticipating a market failure that occurs when market demand is driven by random shocks: during extreme events there might be not enough supply to serve all the consumers. In the bridge example, the supply, i.e. the carrying capacity of the bridge, was not broad enough to serve all the citizens in the extreme event of a catastrophe. A risk that, in fact, affected all the islanders at the same time. In technical words, the risk of scarcity during an extreme event was highly correlated among consumers.

The idea behind capacity markets is similar to the idea proposed by the island’s president for a new bridge. In a quantity-based capacity market, the regulator fixes a certain amount of capacity – which is typically equal to the forecasted peak demand plus a reserve margin – and generators}

A well-designed capacity market allows consumers to enjoy a more reliable system in which brownouts and blackouts are less likely to occur.

compete to achieve this level of capacity that consumers, typically through electricity retailers, may require at some point in the future. This competition between generators to supply future capacity commonly takes place through bidding in centralized auctions (as is the case, for instance, in the PJM market) but it may also result from decentralized bilateral agreements between generators and electricity retailers (for instance, in the California ISO). With the funds they obtain, the electricity generators are able to make new investments and guarantee a certain level of installed capacity, i.e. guarantee that the availability of resources is adequate to meet future demand.

Given the stochastic (random) nature of electricity demand, setting a high enough level of supply to serve the maximum forecasted peak demand reduces extreme events. The extreme event in our tale was the unexpected catastrophe. In power markets, an extreme event – for instance, an extremely hot summer day – eventually may lead to brownouts and rolling blackouts, as the ones occurred in California in 2000 and 2001 or in the Northeastern and Midwestern United States in 2003.

However, just as those islanders would have had to pay a higher toll if the new bridge were to be built, in the electricity sector, current consumers ultimately pay the compensation that electricity generators receive for building additional units of capacity. Therefore, as Kleit and Michaels (2013a) state, capacity markets increase the price of electricity for consumers. However, in return, a well-designed capacity market allows consumers to enjoy a more reliable system in which brownouts and blackouts are less likely to occur.

III. Price Caps and the ‘Missing Money’ Problem

According to the most basic economic principles, in
liberalized and deregulated markets supply and demand are adjusted via market-clearing prices. Then why doesn’t the market provide enough generation capacity (supply) at a higher price for consumers during scarcity events? The answer to this question is given by another risk that is better understood by policymakers: price risk. During scarcity events, electricity generators are willing to skyrocket wholesale prices. As a result, consumers suffer a huge increase in their electricity bills after a scarcity event. Given the unpopularity of huge increases in electricity bills, many regulators have set price caps in wholesale power markets.

Price caps – or ceiling prices – are administrative actions imposed in the spot electricity market limiting the maximum market price of electricity during scarcity periods. They are present in the vast majority of electricity markets, and it is usually argued that they have been put in place to reduce market power, although a major purpose in practice may be to limit politically unpopular price-spikes. 

However, according to Hogan et al. (2005), caps on prices in restructured and liberalized markets have reduced the payments to electricity generators – especially to the peaking plants – that could be used to fund investment in new plants or simply to cover their operating costs.

Therefore, price caps ultimately lead to the so-called revenue adequacy problem or “missing money” problem. The missing money problem occurs when the system fails to attract investments in capacity generation sufficient to meet system operators’ or policymakers’ reliability objectives, as described in Pfeifenberger et al. (2013). More formally, Joskow (2013) states that the missing money problem arises when “the expected net revenues from sales of energy and ancillary services at market prices provide inadequate incentives for merchant investors in new generating capacity.”

Furthermore, in the same paper Joskow (2013) emphasizes that capacity underinvestment has emerged “as a problem in many organized wholesale electricity markets and has been of growing concern in liberalized electricity markets in the U.S. and Europe.” In other words, the missing money problem has become as common among deregulated markets as price caps. Thus, by imposing price caps, regulators have avoided politically unpopular price-spikes only at the cost of creating the missing money problem, which eventually leads to capacity shortages.

IV. The ‘Missing Insurance’ Problem

As we have just argued, a deregulated market with price caps leads to the problem that is illustrated in the island tale. That is, in the case of an extreme event, there might not be enough supply to serve all the consumers in the market.

This market failure is sometimes referred to as the “missing insurance” problem. The missing insurance problem arises when there is a risk that the market is not able to solve and, therefore, is born by the demand side (consumers). In the power market case, given the stochastic nature of demand, the “uninsured” risk for consumers is, for instance, the risk associated with extreme weather conditions. On an extremely hot summer day, households are likely to turn on the A/C. If the demand for electricity exceeds the generation capacity in the system (i.e. there is a scarcity event), some consumers will not be served and, therefore, will suffer blackouts.

As in the island tale, if all consumers were well coordinated in such a way that some of them consume electricity just in the morning and others consume electricity just in the afternoon,
there would be no over-demand, nor scarcity events. However, that is not usually the case. In fact, the problem of power markets is precisely the opposite: consumption is highly correlated. In other words, all consumers are likely to consume electricity around 5:30 p.m., on hot days during summer and on cold days during winter (peak demand periods). On the contrary, households are also less likely to consume electricity in the middle of the night and at early hours of the morning during mild seasons (off-peak demand periods). The correlation in consumption implies that the risk of blackouts due to insufficient generation capacity is also likely too affect too many consumers at once. In other words, the correlation in consumption implies that the risk of blackout is also highly correlated among consumers. This feature exacerbates the missing insurance problem. In fact, blackouts, when occur, are “mass” problems with widespread consequences – as the California, Northeastern, and Midwestern experiences in the U.S. tell us.

V. Solving the Volatility-Reliability Consumer Welfare Trade-off

Following the arguments above, the final trade-off can be stated as follows. On the one hand, a market without price caps is able to provide adequate generation capacity to cover demand even during scarcity events, at the cost of leaving the consumer more subject to (unpopular) price spikes.9 On the other hand, the introduction of price caps leads to less investment in capacity (i.e. the missing money problem), which implies a greater probability of a blackout for consumers.10 However, as argued above, a market with price caps is also able to reduce the risk of electricity price spikes for consumers.

Putting aside the question of whether capacity compensations should be allocated using one mechanism or another (e.g. via auction or via bilateral trading) the introduction of a well-designed capacity market has two effects.

First, it increases the amount of investment. Since the regulator sets a mandatory amount of capacity equal to the forecasted peak demand plus a reserve margin, consumers are guaranteed that, even in the worst forecasted event, there will be enough supply to satisfy all the electricity demand. As a consequence, the risk of capacity shortages disappears.

Second, the introduction of a capacity market together with the presence of price caps is also able to mitigate the unpopular problem of price spikes without creating the underinvestment problem or missing money problem (which is solved, precisely, because of the presence of a capacity market, which imposes a mandatory target level of capacity). As a consequence, the risk of price spikes also disappears.

In summary, a capacity market reduces the blackout probability and decreases the price spike risk relative to a market without a price cap. It does so, however, at the cost of increasing consumer prices. This is because, ultimately, electricity consumers pay the compensation that generators receive to build additional capacity, as explained by Hunt (2012), Thomas et al. (2014), and Arizu et al. (2004), among others.

In the presence of a capacity market – and price caps – consumers pay higher electricity prices but avoid the risk of capacity shortages (i.e. blackouts) and the risk of price spikes. Therefore, a capacity market solves the missing insurance problem and serves as an insurance mechanism for the consumers.
VI. A Case Study: The Texas ERCOT Market

The Electric Reliability Council of Texas (ERCOT), created in the 1970s and established in 1996 as the first independent system operator (ISO) in the U.S., manages 85 percent of Texas’ electric load. In 2011 it served around 23 million customers, and had a total installed capacity of 84,000 MW.

Texas ERCOT has traditionally relied on an “energy-only” market system. In other words, wholesale electricity prices adjust to equate demand and supply. If there is a shortage, prices rise above the marginal production cost and generate a scarcity rent. This scarcity rent is needed to recover investment costs. Capacity compensation payments are absent in the system.

Although the ERCOT market appeared to be a successful case study of a deregulated energy-only market, in 2010 and 2011 several reports warned that the system was not providing the right incentives to build new generation capacity. In fact, as reflected in Figure 1 the capacity reserve margin, far from reaching the 13.75 percent target level, has been shrinking and is projected to become negative.

In an attempt to solve this problem, in 2012 policymakers raised the so-called System-Wide Offer Cap (SW-CAP), which is the highest per-MWh price at which a generation resource can offer wholesale energy into the market (i.e. the price cap). By doing so the regulator expected to generate greater scarcity rents and, consequently, to incentivize new investment in generation capacity. Thus, as shown in Table 1, the price cap in ERCOT experienced a steady increase from $2,500/MWh in 2011 to $9,000/MWh in 2015. In fact, ERCOT announced a long-term goal for the price cap of $15,000/MWh. Thus, from 2011 on, the ERCOT wholesale price cap has been set well above the price cap in other markets such that the NYISO ($1,000/MWh), PJM ($1,000/MWh), or ISONE ($1,000/MWh).

However, this created a perverse incentive for suppliers. As we explained above, the absence of price caps (or, similarly, abnormally high price caps) skyrocket consumers prices during scarcity events. That was precisely the case on Sept. 3, 2013. As shown in Figure 1, prices increased by almost 5,000 percent during the afternoon, due to a supply tightening.

Questions continued to be asked about the reliability of the ERCOT system beyond 2013. Thus, in June 2014, ERCOT implemented the so-called Operating Reserve Demand Curve (ORDC) system. The basic idea underlying this mechanism is that generators holding additional reserves are compensated whenever total system reserves cross a lower threshold. The compensation is immediately added to the wholesale price.

As of the time of writing it is still unknown whether this change will solve the problem. In

![Figure 1: ERCOT Forecasted Reserve Margin (2011)](image-url)
any case, it seems that ERCOT members are reluctant to embrace the idea of a capacity market. In fact, although some authors have claimed that capacity markets have successfully solved similar capacity shortages in a range of different countries and regions – for example, see Ausubel and Cramton (2010) – other authors, such as Kleit and Michaels (2013a,b,c), have argued against the implementation of capacity markets in Texas. In fact, the question of whether or not a capacity market would be welfare-enhancing for ERCOT consumers is still open.

VII. Conclusions

With this article we have sought to enrich the current debate about price caps and capacity markets. We have focused on the consequences of capacity markets for final consumers of electricity.

We argue that a capacity market intends to protect the consumers in the sense of avoiding inadequate capacity and hence blackouts or brownouts in future periods by achieving adequate capacity and avoiding scarcity events while having price caps. All these benefits come at the cost of making consumers pay a compensation that will make electricity prices more expensive (but less volatile) on average. In a sense, as an insurance contract prevents its underwriter from some risk at the cost of paying a premium, a capacity market serves as a solution for the “missing insurance” problem in power markets.

References


Endnotes:

1. According to Creti and Fabra (2007), there are two main types of capacity payment systems: price-based systems and quantity-based systems. The price-based mechanisms have barely been analyzed in the past – see, for instance, Newbery (1995) and Wolak and Patrick (2001) – and have been excluded from the current debate. Thus we focus our analysis on the quantity-based systems, which are actually the ones that the aforementioned countries and regions are considering to introduce. A broader classification is in Spees et al. (2013); they include the so-called price-based system in the group of “Administrative mechanisms,” while the quantity-based systems are included in the group of “Market-based mechanisms.” Furthermore, they distinguish between the “LSE RA requirement systems,” in which the capacity allocations are achieved through bilateral agreements (e.g. CAISO); and the (proper) “Capacity Markets,” in which centralized auctions are used (e.g. PJM, NE-ISO, NYISO).

2. Historically, the target that guarantees adequate capacity has been set in such a way that the system is expected to experience a supply shortage once every ten years. As Spees et al. (2013) point out, this is usually called the 1-event-in-10-years loss of load expectation (LOLE).

3. As Hogan (2013) points out, the product of these markets is installed capacity, not energy.

4. A deep analysis of this crisis is by Weare (2003).


6. Typically retailers pay such compensations; however, as several authors support – Hunt (2012), Thomas et al. (2014) and Arizu et al. (2004) among others – these compensations are effectively passed through to final consumers (via higher electricity prices).


8. As Hartley (2014) points out, in some markets such as Texas ERCOT and MISO, prices in the spot market have often been negative during off-peak demand hours. That is, generators are willing to pay buyers to consume power!

9. See, for instance, the Texas ERCOT experience on September 3, 2013 in the following section.

10. See Hogan et al. (2005).

11. The commonly employed term ‘‘energy-only’’ market captures the situation in which the only mechanism used in the power sector is a wholesale market, with no capacity payments of any kind.

12. Source: Texas ERCOT.

13. The capacity reserve margin is calculated as follows: (Generation Capacity Resources (MW) – Firm Load Forecast (MW))/(Firm Load Forecast (MW)) * 100.
