Ten propositions on electricity market design:

Energy-only vs. capacity markets

Lion Hirth[1] & Falko Ueckerdt[#]

a) Introduction

In several European countries, market participants, system operators and policy makers debate the "design" of wholesale electricity market. Specifically, it is discussed if energy-only markets provide sufficient incentives for generation investments, and if the introduction of "capacity markets" is needed. This article recaps fundamental arguments from the economic literature on the topic, some of which is more than 60 years old. In turns out that there is little in theory that suggests that capacity markets are needed.

The UK started liberalizing its electricity market around 1990, and many European countries followed in the subsequent decade. During this process, a wholesale "market design" emerged where most transactions are denominated in terms of energy (€/MWh). For example, European power exchanges trade electricity in energy terms only. This setup has been labelled an "energy-only market".[2]

When the macroeconomic crisis hit the electricity sector in late 2008, profit margins started to drop. Since 2009, and for the years to come, short-term profits are significantly below the fixed costs of power plants, implying that generators currently do not earn their expected rate of return, and that there is no inventive to invest in new power plants (Figure 1).

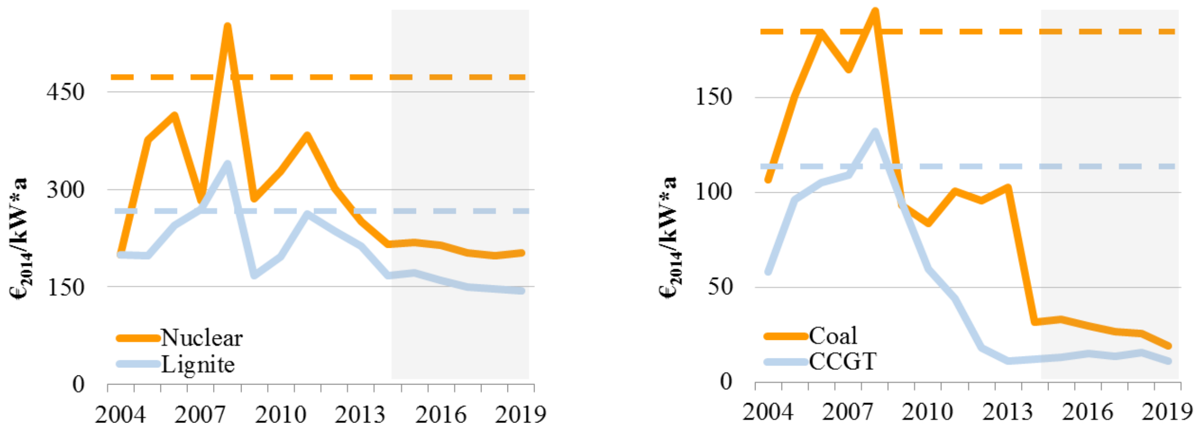

Figure 1. Gross margins of thermal plants in Germany (bold lines), based on historical spot prices and futures. Margins have dramatically fallen since 2008 and are significantly below the level that is needed to trigger new investments (dotted lines).

During the last years, the discussion on investments incentives has become quite political. Many observers question that energy-only markets provide sufficient incentives for generation investments, often pointing at low prices and spreads. It is discussed if "capacity markets", also called capacity payments, capacity mechanisms, or capacity remuneration schemes, need to be introduced to ensure resource adequacy. In such systems, capacity (MW) are remunerated, usually on top on energy prices. Today (early 2014), several European countries have introduced some sort of capacity remuneration scheme[3] and in some large markets such as the UK, France, and Germany it is discussed to introduce extensive capacity markets.

Several recent studies suggest that some sort of "capacity mechanism" is warranted to incentivize generation capacity in Germany and that energy-only markets cannot provide the corresponding incentives, such as Cramton & Ockenfels (2011) and EWI (2012). Ecofys (2012) is more cautious. Hogen (2005) provides a good documentation about the (quite similar) discussion in the US a decade ago.

We have the impression that in the public, some arguments are poorly understood and some facts are misinterpreted. This paper recaps fundamental arguments from economic theory and some basic insights from power system engineering in the form of twelve propositions. It seems that the arguments for capacity remuneration

[1] Vattenfall GmbH & Potsdam-Institute for Climate Impact Research, hirth@pik-potsdam.de, +49 30 8182 4032.
[#] Potsdam-Institute for Climate Impact Research, ueckerdt@pik-potsdam.de, +49 331 288 2067.
The findings, interpretations, and conclusions expressed herein are mine and do not necessarily reflect the views of Vattenfall or PIK.
[2] In most energy-only markets, bilateral contracts can be signed between market participants, and these contracts can include all kind of payments, including fixed and capacity payments.
[3] In Germany, the *Reservekraftwerksverordnung* authorizes the regulator to contract capacity under the so-called "winter reserve". Similar tools, often called a "strategic reserve", are currently implemented in Poland, Belgium, and Great Britain.

schemes rest of weak economic foundations: under a quite broad set of assumptions, energy-only markets pride efficient signals for generation investments.

The remainder of the paper consists of three sections. a) We argue that energy-only markets work under a wide set of assumptions. b) We explain where they are potentially subject to deficits. However, capacity markets can be subject to the same (and additional) deficits. c) We then comment on the arguments put forward in the European debate, with a focus on Germany. d) Some basic thoughts on the potential design of a capacity remuneration scheme are presented.

b) How (and why) energy-only markets work

1. Energy-only markets with scarcity pricing provide efficient price signals for generation investments.

It is sometimes believed that the equilibrium electricity price in energy-only markets always equals the variable cost of the marginal plant. This is *not* the case. Instead, prices rise above variable costs of generation if and when overall generation (and import) capacity is scarce. Such prices are called "scarcity prices".

In energy-only markets, there exists a long-term equilibrium, where all generators, including the "last generator", earn their capital costs. In this equilibrium, both the resulting overall amount of capacity and the technology mix is welfare-optimal. In fact, these equilibrium prices reflect not only variable cost, but also the opportunity cost of capacity. Hence, although prices are always expressed in terms of energy (€/MWh), the term "energy-only market" is somewhat misleading.

There is a long tradition of energy economics literature that shows and proves these findings theoretically (Boiteux 1951, Steiner 1957), reviewed by Crew et al. (1995).

2. This holds under a wide set of assumptions.

Proposition 1 holds under a fairly wide set of assumptions: unlike some observers suggest, a long-term equilibrium exists for price-elastic as well as perfectly price-inelastic demand. If demand is very price-inelastic, the equilibrium price raises very high during very short time spans. If demand is more price-elastic, price spikes are lower and distributed but with longer duration.

Proposition 1 also holds with (high) shares of variable renewables, and with (high) shares of generators with very low variable costs. With large amounts of wind and solar power, prices will be subject to a higher variability.

It is a myth that energy-only markets can only work with elastic demand, or can only work with low shares or wind and solar power. There exists not "renewables threshold" above which they will stop working.

3. Under scarcity pricing, economic agents are exposed to price risk. This is good.

The level of peak load is uncertain in advance. In energy-only markets, this is translated into uncertainty about price peaks. An investor cannot be sure if and how frequent scarcity prices will occur.

Some argue that uncertainty renders scarcity pricing infeasible. However, price risk is not induced by market design, but a fundamental property of the real world. It is inherently uncertain how demand (and other parameters) will develop. Under any market design, peak capacity might be unused. It is efficient that producers and consumers bear this risks, not society or the state.

4. If capacity is insufficient, load will be shed. Load shedding is not a blackout.

A "blackout" is a widespread outage of the transmission system. A blackout can have severe consequences because of its cascading nature, its long duration, and because most power plants cannot be restarted without electricity from the grid.

If investors are underestimating electricity consumption, and demand is very price-inelastic, it can happen that energy-only markets do not clear. In this situation, generation capacity is insufficient. It is sometimes argued that capacity shortage causes a blackout, which, because of its dramatic consequences, needs to be avoided at any cost. However, this is not the case. Lack of capacity does not cause a blackout.

If markets do not clear, system operators have a multitude of possibilities to react on. If day-ahead markets do not clear, around 30 hours remain to take these measures, given that scarcity usually occurs in the evening hours. System operators shed individual industrial consumers or individual distribution grids (ahead of time), and activate balancing power (in real time). If imbalances remain, *certain* loads will be shed in real time automatically by frequency relays ("brownout"). These measures can be quite costly for society, but not nearly as costly as a blackout.

There is a widely cited theoretical paper by Joskow & Tirole (2007) that argues that there exists a supply externality in power markets. Their major results rests essentially on the argument that whenever capacity is scares, blackouts occur. As we have explained, this is not the case.

c) Energy-only markets have some weaknesses

5. Energy-only markets are exposed to regulatory risk. (But capacity markets might be even more exposed.)

Scarcity pricing works only if scarcity prices are allowed to occur. If regulators prevent price peaks (or investors anticipate this might happen), they can obviously not sufficiently incentivize investments.

System operators and regulators have ad-hoc possibilities to lower prices, such as tampering with the amount of balancing power to be reserved, (threaten to) punish suspected abuses of market power, or command and control measures similar to those recently implemented in Germany.[4] This induces political risk to scarcity prices, potentially leading to inefficiencies. However, this is not a failure of markets, but of regulation.

Regulatory risk is not specific to electricity or energy-only markets: all markets are prone to regulatory risk – arguably some more than others. We believe it is probable that introducing capacity markets might increase regulatory risk instead of reducing it. The first-best solution to regulatory risk is not a capacity market, but to think of ways to increase policy commitment. For example, power market regulators could work independently from government, similar to central banks.

6. There is potential for abuse of market power during scarcity events.

If capacity is scarce, generators have the possibility to exercise market power. It might be difficult to distinguish competitive scarcity prices from market power abuse.

However, in the long-term equilibrium, market power abuse is inconsistent with market entry. Competitive pricing is the only equilibrium solution.

Hogan (2005) proposes to target capacity withholding instead of capping prices as a way to ensure competitive bidding while allowing scarcity prices to exist.

7. If investors are more risk-averse than society, energy-only markets can be inefficient.

If a number of conditions coincide, energy-only markets can cause underinvestment in generation capacity: investors are disproportionally risk-averse; peak demand is uncertain; and demand is very price-inelastic.

Higher or more frequent scarcity prices would be needed to compensate investors for their risk-averseness. Markets would still clear, no blackouts would occur, but prices would be inefficiently high.

In this case, the market failure is diverging risk aversion between investors and society. Only if society is less risk averse than investors, a planner would invest more than market actors.

In any case, to us it remains unclear if capacity markets are an efficient instrument to compensate for this.

d) Remarks on the current policy debate in Europe

8. The low level of wholesale prices indicates that markets work rather than that they fail.

Wholesale prices are currently low in Germany, as anywhere in Europe, because capacity is abundant. This is a sign of markets at work, not one of market failing. Low prices do not require public intervention!

Four major factors have depressed European power markets: capacity inherited from times before the liberalization (especially long-living nuclear plants), technological progress (market coupling), policies (subsidizing renewables), and the macroeconomic recession (that surprised utilities).

Low prices are the signal to investors that prevent them from building plants that are not required. The reason for low prices is that new capacity is *not* needed. In fact, it seems quite absurd to argue low prices indicate that capacity should receive additional remuneration.

Once capacity will become scarce, forward prices and price projections will indicate this several years ahead of time.

9. Capacity scarcity in Germany is a locational issue. There are better policies than a capacity market to solve these issues.

It is not overall capacity that is scarce, but regional capacity in South-West Germany and Northern Bavaria. Resource availability (wind power, lignite) and fuel prices (hard coal) favor generation investments in the North and Northeast of Germany, and transmission capacity is constrained. Hence the proposals for capacity mechanisms look in the wrong direction: It is not capacity that is scarce, but capacity *at certain locations*.

Under current market design regional scarcity is not reflected in prices – because there is only one single price for Germany.

---

[4] Since 2012, the German regulator can force utilities to keep power plants in operations if the owner decides to decommission the asset. Since 2013, this is regulated in the *Reservekraftwerksverordnung*.

Transmission congestion can be solved by grid investments or by locationally differentiated energy-only price signals, such as bidding areas or nodal pricing. Logically, an energy-only market that lacks location price signals cannot provide locationally efficient investment signals. The absence of locational price signals can be understood as incomplete market design. A capacity market can never efficiently solve locational constraints.

10. The degree of demand elasticity is not known, but could be significant in the medium term.

More elastic demand would ease some issues with energy-only markets, such as market power abuse and uncertainty about price peaks. The strong statements on demand inflexibility by Cramton & Ockenfels (2011) are unproved hypotheses. Up to know, demand behaved quite price-inelastic – because prices have never been high!

Industrial loads (40% of German demand, of which half is in energy-intensive industries) might be actually quite elastic at high prices. This empirical question cannot be answered with historical data, because never since the liberalization there was a period of persistently high prices: since 2001, only during three hours the wholesale price has exceeded 1000 €/MWh. In U.S. power systems, interruptible load is up to 11% of peak load, and demand heavily bids into capacity markets (Hurley et al. 2013).

Once high prices occur more often, price signals are likely to induce consumers to develop ways to become more price-sensitive through all kinds of innovation and technological change.

e) Thoughts on design principles of a "capacity market"

While we believe economic theory does not support the need of a capacity remuneration system, the political reality renders its introduction quite likely in a number of European countries. If such a scheme is to be introduced it should be designed in a way that it interferes with the energy market as little as possible. Most importantly, capacity that is remunerated should be dispatched only of the market does not clear, as a *last resort*. It might be called an "emergency reserve". The emergency reserve is an instrument to ensure generation capacity adequacy.

As in many other policy areas, including monetary policy, such an instrument is subject to severe commitment problems. The weaker the commitment, the more the price signals for investments is blurred. For credible commitment, the dispatch decision should be transferred to a legally and economically independent institution, similar to a central bank – a "capacity bank". The capacity bank should not only be independent from government, system operators, power exchanges, and market actors, but also from the regulator. Furthermore, the dispatch decision should be rule-based: the capacity bank is only allowed to dispatch the emergency reserve if the day-ahead market does not clear despite a very price (say, 10,000 €/MWh), and if the intraday-market does not clear either.

The central bank should assess the amount of capacity that is needed and procure the reserve several years ahead of time in public tenders. Selection should be based on one criterion only: an availability-adjusted capacity price. An emergency reserve should not be used to pursue ends other than securing generation adequacy, such as climate policy, renewable expansion, or ensuring uniform bidding areas.

References

Cramton, Peter & Axel Ockenfels (2011): *Economics and design of capacity markets for the power sector*, report for RWE.

Energiewirtschaftliches Institut (2012): *Untersuchungen zu einem zukunftsfähigem Strommarktdesign*, report for Bundesministerium für Wirtschaft und Technologie.

Ecofys (2012): *Notwendigkeit von Kapazitätsmechanismen*, report for Bundesverband der Energie- und Wasserwirtschaft.

Michael Crew, Chitru Fernando, Paul Kleindorfer, 1995, A theory of peak load pricing: A survey. Journal of regulatory economics.

Boiteux, M. 1960, Peak load pricing. Journal of business (French original published in 1951)

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

Steiner, Peter (1957): "Peak Loads and Efficient Pricing", *Quarterly Journal of Economics* 71(4), 585 – 610.

Hogan, William (2005): "On an "Energy Only" Electricity Market Design for Resource Adequacy", *Working Paper*.