EPRG WORKING PAPER

A Primer on Capacity Mechanisms

EPRG Working Paper   1806

Cambridge Working Paper in Economics   1814

Natalia Fabra

Abstract   A simple model is built up to capture the key drivers of investment and pricing incentives in electricity markets. The focus is put on the interaction between market power and investment incentives, and the trade-off it introduces when designing the optimal regulatory instruments. In contrast to the energy-only market paradigm that assumes perfect competition, our model demonstrates that in the presence of market power scarcity prices do not promote efficient investments, even among risk-neutral investors. Combining price caps and capacity payments allows to disentangle the two-fold objective of inducing the right investment incentives while mitigating market power. Bundling capacity payments with financial obligations further mitigates market power as long as strike prices are set sufficiently close to marginal costs.

Keywords   scarcity pricing, market power, capacity markets, reliability options.

JEL Classification   L13, L51, L94

A Primer on Capacity Mechanisms

Natalia Fabra[*]

Universidad Carlos III de Madrid and CEPR

February 12, 2018

Abstract

A simple model is built up to capture the key drivers of investment and pricing incentives in electricity markets. The focus is put on the interaction between market power and investment incentives, and the trade-off it introduces when designing the optimal regulatory instruments. In contrast to the energy-only market paradigm that assumes perfect competition, our model demonstrates that in the presence of market power scarcity prices do not promote efficient investments, even among risk-neutral investors. Combining price caps and capacity payments allows to disentangle the two-fold objective of inducing the right investment incentives while mitigating market power. Bundling capacity payments with financial obligations further mitigates market power as long as strike prices are set sufficiently close to marginal costs.

Keywords: scarcity pricing, market power, capacity markets, reliability options.
JEL Codes: L13, L51, L94.

1 Introduction

World-wide, the electricity sector is going through fundamental changes triggered by the massive deployment of renewables. One of the key challenges in renewables-dominated systems is to ensure security of supply at all times. The intermittency of renewable resources, coupled with the weakness of demand response, requires that conventional technologies must be available to ensure that there is always enough capacity to meet demand. Yet, renewables depress wholesale electricity prices and increase their volatility, which in turn reduces the profitability of investments in back-up capacity.

In this context, there are growing concerns about the inability of the current market and regulatory arrangements in electricity markets to induce adequate investments in generation capacity. Essentially, there are two confronted views. On the one hand, the so-called energy-only market paradigm (Hogan, 2005) advocates for the removal of price caps, as these preclude firms from obtaining the scarcity rents that would otherwise cover the fixed costs of their investments. According to this view, price caps are thus at the core of the *missing money problem* that creates under-investment.[1] The alternative view is that generation capacity has an intrinsic value that is distinct from the sale of energy, i.e., adding new capacity improves reliability even when it is not actually used to produce. The failure to price such externality is thus the root of under-investment. This view has prompted several countries to supplement firms' energy market revenues with capacity payments in order to strengthen their incentives to invest in generation capacity (European Commission, 2016).[2]

In this paper, I build a simple model to shed light on some of the questions at the center of the regulatory debate about the need, effect and design of capacity payments in the electricity sector. In particular, the model is used to address the following questions: can we rely on scarcity pricing to promote efficient investments in generation capacity, or should we rather combine price caps and capacity payments to ensure security of supply at least cost? What is the impact of capacity payments on the performance of energy markets? How much capacity should be procured and how does this depend on the level of the price cap or the degree of market power? Does it matter whether all plants, or only the new ones, receive capacity payments? What are the effects of market power in the capacity market and how could these be mitigated through market design? Finally, should capacity payments be bundled with financial commitments, and if so, which are the optimal ones?

A common thread of this paper is the idea that investment incentives and firms' ability to exercise market power cannot be analyzed in isolation. Understanding their interaction is key both when diagnosing the source of market failures, as well as when designing the regulatory instruments needed to address them. The energy-only market paradigm, which has been highly influential in the policy arena, advocates for the removal of price caps as a way to restore investment incentives

---

[1]Joskow (2007) provides numerical examples to illustrate the missing money problem.

[2]This paper will not assess the option of promoting demand side response and storage in order to bring in the flexibility needed to cope with intermittent resources. Demand side response requires the mass deployment of smart meters as well as the change in pricing rules towards Real Time Pricing. Similarly, the high cost of storage rules it out as a viable alternative in the short run.

with no need to resort to capacity payments. However, this is a direct consequence of ruling out market power by assumption: in perfectly competitive markets, price caps simply have no role to play in mitigating market power. In contrast, by endogenously allowing for market power, this paper shows that the removal of price caps triggers capacity expansions but it does so at the cost of strengthening market power. Accordingly, scarcity pricing is not an efficient solution for promoting investment.

Providing adequate investment incentives while at the same time mitigating market power requires the use of capacity payments in conjunction with price caps since price caps alone would mitigate market power but would result in poor investment incentives. As Cramton and Stoft (2006) put it, capacity payments are motivated by the goal that "the missing money must be restored without reintroducing the market power problems currently controlled by price suppression." In line with this view, this paper analyzes the role of capacity payments in imperfectly competitive markets.[3] Purposely, it omits other ingredients (such a investors' risk aversion, or the social costs of lost load) that would strengthen the main conclusions of the paper.

Capacity markets are one commonly used instrument to determine capacity payments. The regulator (or the System Operator) sets the volume of capacity to be procured, investors submit the prices at which they are willing to make their capacities available, and the capacity price is set through market clearing. If the capacity market is competitive, the capacity price will cover the investment costs net of the scarcity rents that firms receive in the energy market. This links capacity payments with the level of the price cap: the more stringent it is, the lower the scarcity rents and the higher the resulting capacity payments. Despite the increase in capacity payments, a stringent price cap policy saves consumers more than it costs as it is more effective in reducing market power rents.

This conclusion must be qualified if there is market power in the capacity market. First, market power in the capacity market makes it more costly for the regulator to procure a given amount of capacity (as firms retain the market power rents that arise both in the energy market as well as in the capacity market); and second, this might in turn induce the regulator to procure less capacity than would otherwise be optimal. Hence, continued efforts must be devoted to preventing market power in capacity markets, both through good market design as well as through close market supervision.

In the same vein, capacity payments are more efficient when paid to the new plants only (i.e., so-called *targeted* mechanisms): for a given capacity price, aggregate investment is the same as when all plants receive capacity payments (*market-wide* mechanisms), but since it is less costly to induce investment under targeted mechanisms, the regulator optimally decides to procure more capacity. Thus, in equilibrium, aggregate investment is closer to the first-best when only new plants receive capacity payments, making consumers better off. Paying for the old plant's capacity would

---

[3]To be sure, there are other ways to address the investment problem beyond capacity payments. For instance, auctions for new investments that determine the price per MWh produced by the new assets would also serve that purpose. This paper focuses on capacity payments because of the attention they currently receive in the regulatory debate.

over-compensate firms beyond their lost profits.[4]

Another controversial aspect of capacity payments is that they do not incentivize firms for making their capacities available other than through explicit penalties. For instance, the European Commission (2016) reports that in many instances, firms receiving capacity payments have limited obligations; in turn, since the penalties for non-availability are low, firms face insufficient incentives for making their plants reliable. This is one of the reasons why regulators are increasingly resorting to so-called *reliability options*, which embody an endogenous penalty for not being available. Since such a penalty is more costly the higher the market price, firms have stronger incentives for being available during periods of scarcity. Additionally, reliability options help mitigate market power in the energy market.

Similar to capacity markets, reliability options involve quantity regulation. However, the two regulatory solutions differ in two key aspects. First, reliability options allow for plant-specific price caps, in contrast to capacity markets that rely on market-wide price caps. Thus, auctioning reliability options is a more effective tool for preventing market power, particularly in the presence of several generation technologies. Furthermore, whereas reliability options are backed by a contract, market-wide price caps and capacity payments are subject to greater regulatory uncertainty. As Joskow (2007) puts it: "Market rules and market institutions change so frequently, that the opportunities for regulators to hold-up incumbents by imposing new market or regulatory constraints on market prices are so great that uncertainty about future government policies acts as a deterrent to new investment."

Regulators play an important task when designing reliability options, as their potential for preventing market power depends on how close the strike price is to the plants' marginal costs. A precise estimation of marginal costs is not always feasible, but it is not indispensable either. Energy regulators have good knowledge of the determinants of marginal costs and these can serve to set strike prices *reasonably* close to marginal costs. Setting strike prices that are too high would lead to reliability options rarely being exercised, and firms would obtain capacity payments with only weak disciplining effects. In turn, this has implications regarding the suitability of technology-neutral auctions for reliability options: if the strike price is set high enough so as to make room for even the most expensive technologies (e.g., demand response at the VOLL), the disciplining effects of reliability options on the lower cost technologies would be undermined.

Much has been written about the need, effect and design of capacity mechanisms. However, as far as I am aware of, there have not been previous attempts to analyze the various issues within a single analytical framework. Cramton and Soft (2006), Joskow (2007), and Bushnell *et al.* (2017) provide excellent discussions of several of the issues covered here, but their aim is not to formalize them. Other papers build models to shed light on specific topics, with two issues attracting most of the attention: the impact of capacity payments in interconnected systems, and the need of capacity payments following an increase in renewables. For the former, see Fabra and Creti (2006), Crampes

---

[4]In the UK, the rationale for using a market-wide mechanism was to avoid inefficient exit. This issue is omitted from the analysis of this paper.

and Salant (2018) and Lambin and Leautier (2018);[5] for the latter, see Llobet and Padilla (2018). Additionally, some papers have analyzed investment incentives in energy markets but with few exceptions (Bajo-Buenestado, 2017; Fabra *et al.* 2011; Zottl, 2011), most of the existing models abstract from market power issues.

The paper is structured as follows. Section 2 describes the basic model of capacity investments in electricity markets, and Section 3 enriches it by adding capacity payments. Section 4 explores further issues, such as the comparison of targeted versus market-wide mechanisms, the impact of market power in the capacity market, or the design and effects of reliability options. Section 5 of the paper concludes. The appendixes contain an analysis of extensions, as well as the proofs for the main propositions.

2   Model Description

Consider a market in which n firms compete to generate electricity. All firms have the same production technology, with marginal costs normalized to zero up to the firm's capacity, k_i for $i = 1, ..., n$, and per-unit capacity investment costs $c > 0$. Production above capacity is infinitely costly, and firms' capacities are always available.[6]

Demand in this market, denoted θ, is assumed to be perfectly inelastic up to consumers' reservation value, v (in the industry's jargon, v is the Value of Lost Load or VOLL).[7] The demand parameter θ is uniformly distributed in the unit interval. Realized demand is known when firms take their production decisions, but unknown at the investment stage. The implicit assumption is that investment is a long-run decision while production is a short-run decision: the latter is taken on a daily or hourly basis when firms have precise estimates of demand, whereas the former is taken every 20-30 years, with capacity assets facing significant demand variation over their lifetime. Last, to avoid price spikes, the regulator might decide to introduce a price cap, denoted P. Setting $P = v$ is thus equivalent to not having a price cap.

We consider a two-stage game with the following timing. First, firms take simultaneous capacity choices k_i, $i = 1, .., n$, while facing demand uncertainty. Once chosen, capacities become publicly observed. Second, demand θ is realized and observed by all firms. Firms submit simultaneous price offers (or bids) to the wholesale energy market. In particular, a bid $b_i \in [0, P]$ commits firm i to produce up to capacity whenever the market price is at least b_i. The auctioneer dispatches firms' capacities in increasing bid order until all demand is satisfied. In case of a tie in prices, the auctioneer randomly decides which firms among those that bid at the market clearing price are

---

[5]See also Newbery (2016) for a policy discussion.

[6]We are dealing with thermal technologies. In the extensions we allow for renewable capacities, which are intermittent and thus not always available. See also Llobet and Padilla (2018). Zottl (2011) allows for investments in a base-load and a peak-load technology.

[7]We are assuming inelastic demand because that is the case in practice. The deployment of smart meters and the use of real-time pricing might alleviate this. Still, capacity payments will be needed whenever electricity market prices are capped implying that with some probability the market does not clear, as we show below. Hence, the main conclusions of the analysis would remain unchanged if we allowed for elastic demand.

dispatched. The market clearing price (i.e., the highest accepted bid) is paid to all the dispatched production.

Before characterizing equilibrium investments, we first explore two benchmarks: the First Best capacity, and the capacity choices that arise under free entry and no market power.

2.1 First Best capacity

Defining total welfare W as the sum of consumer and producer surplus, total welfare at the investment stage equals consumers' gross utility from electricity consumption, minus production and investment costs,[8]

$$W = v \int_0^K \theta d\theta + v \int_K^1 K d\theta - cK. \tag{1}$$

Expression above is a function of K only: with price inelastic demand, prices only affect the distribution of surplus between consumers and producers, with no impact on total welfare. Note that we are implicitly assuming that demand can be rationed in case of shortages (i.e., there is no involuntary rationing, or in the industry jargon, the System Operator can implement rolling blackouts so as to avoid a system collapse when demand exceeds aggregate capacity).[9]

Maximization of (1) with respect to K implies that the optimal or First Best (FB) aggregate capacity is given by:

$$\begin{aligned} \frac{\partial W}{\partial K} &= v\left[1 - K\right] - c = 0 \\ &\Rightarrow K^{FB} = \frac{v - c}{v} < 1. \end{aligned} \tag{2}$$

At First-Best capacity, there is some probability of rationing because the regulator trades-off the cost of giving up consumption (v) versus the fixed cost of extra capacity (c). In other words, fully eliminating rationing is not optimal. In practice, this optimality condition is well known by system operators. For instance, Newbery (2016) reports that the National Grid Company in the UK chose how much capacity to procure by balancing the cost of additional capacity against the cost of expected energy unserved.

2.2 The energy-only market paradigm

The energy-only market paradigm states that uncapped wholesale electricity markets provide the right signals for investment decisions, with no need to resort to capacity payments (see Bushnell *et al.*, 2017). This paradigm rests on two key assumptions: (i) there is free entry; and (ii) there is no market power in the energy market. The former assumption implies that entry/investments take pace until expected profits become zero; the latter assumption implies that prices equal marginal costs whenever there is enough aggregate capacity, but otherwise prices rise up to the price cap

---

[8]Since marginal costs have been normalized to zero, only the latter are explicit in the welfare expression.
[9]Below, we relax this and include a probability with which the system operator fails in avoiding the blackout.

P. Thus, electricity wholesale prices are relatively volatile given that, with inelastic demand, small demand shocks can give rise to large price spikes.

Under these assumptions, the market indeed generates the FB capacity at least cost for consumers if and only if price caps are removed. Focusing on the symmetric equilibrium at which all firms invest k, expected profits for each firm are

$$\pi_i = P \int_{nk}^{1} k d\theta - ck.$$

The first term represents the scarcity rent, i.e., the revenues obtained when prices are increased above marginal costs until demand can be met with available capacity (Borenstein, 2000). In a competitive market with free entry, firms build new capacity as long as the scarcity rent exceeds the cost of capacity. Hence, in equilibrium, each firm invests

$$k^* = \frac{1}{n} \frac{P - c}{P}.$$

Thus, equilibrium aggregate capacity

$$K^* = nk^* = \frac{P - c}{P} \leq K^{FB} = \frac{v - c}{v},$$

is below the first-best, unless $P = v$. In words, to avoid under-investment, price caps have to be removed to allow firms to obtain the full value of scarcity rents.[10] The reason is simple: if $P < v$, the private gain from expanding capacity is below the social gain, and this creates under-investment with respect to the first-best. If prices are capped, scarcity rents are not enough to cover the costs of the optimal investments, leading to what is known as the *missing money problem*. This is the relevant case in practice, as acknowledged by the European Commission (2016): "In the absence of price-responsive demand, rules put in place by national authorities to balance supply and demand often include low regulated price caps that do not reflect customers' willingness to pay for secure supplies and that therefore result in prices which do not reflect the actual value of additional resource adequacy."

Total welfare under the energy-only paradigm is equal to consumer surplus at the first-best capacity (note that firms make no profits because scarcity rents just cover investment costs),

$$W\left(K^{FB}\right) = CS\left(K^{FB}\right) = v \int_{0}^{\frac{v-c}{v}} \theta d\theta. \tag{3}$$

In sum, the energy-only market paradigm concludes that price caps are the source of inefficient

---

[10]If we had allowed for a general demand distribution $G(\theta)$, FB capacity would be $G^{-1}\left(\frac{v-c}{v}\right)$. The result above that scarcity rents equal investment costs at the FB capacity is robust to assuming any general G, as the profit equation $v\left(1 - G\left(K\right)\right)K - cK$ equals zero at the K that satisfies the FOC for FB capacity, $v\left(1 - G\left(K\right)\right) - c = 0$.

investments. It thus suffices to remove price caps to allow the market to deliver optimal investments, with no need to resort to capacity payments.

But, is this conclusion still valid in markets in which the assumptions of the energy-only markets paradigm do not hold? What are the implications of removing price caps in such markets? We turn to these issues next by relaxing two key assumptions in turn: we first introduce involuntary rationing (blackouts) while sticking to the competitive market assumption; we then introduce market power while omitting the possibility of blackouts.

2.3   Involuntary rationing

In the basic model we assumed that demand can be rationed, i.e., during scarcity periods consumption is reduced to meet existing capacity. However, this is not always technically possible in practice. As Joskow (2017) puts it, "individual consumers cannot choose their individual preferred level of reliability when rolling blackouts are called by the System Operator; their lights go off along with their neighbors' light." This is particularly relevant when assessing the market's availability to provide security of supply, as the inability to stop system blackouts when there is not enough capacity implies that capacity investments create a positive externality as they reduce the probability of system blackouts. As this externality cannot be fully reaped by capacity owners, under-investment results. Removing price caps would not fully offset this inefficiency, since the social value of additional capacity exceeds the private gain from consumption.

To shed light on this issue, suppose that in case total demand exceeds available capacity, the regulator cannot stop a system blackout with probability $\gamma < \frac{1}{2}$ (so that consumers cannot consume at all); however, with probability $(1 - \gamma)$ the system does not collapse and there is consumption up to existing capacity.

Total welfare is given by

$$
W = v \int_0^K \theta d\theta + v(1 - \gamma) \int_K^1 K d\theta - cK.
$$

Maximization with respect to total capacity gives the first-best capacity,

$$
K^{FB} = \begin{cases} 1 & \text{if} \quad \gamma \geq \frac{c}{v} \\ \frac{1}{1 - 2\gamma} \frac{v(1-\gamma) - c}{v} & \text{if} \quad \gamma < \frac{c}{v} \end{cases}
$$

The first-best capacity is increasing in γ, i.e., as blackouts become more likely, first-best capacity goes up to avoid them. On the two extremes, if the probability of a blackout is zero ($\gamma = 0$), we recover the same solution as in the basic model; however, if the probability of a blackout is very high, it is optimal to build enough capacity to cover the peak of demand.[11]

---

[11]In electricity markets in practice, regulators require there to be reserve capacity above the peaks of demand. This can be easily accounted for in our model by adding a probability of outage, so that capacity has a probability of not being available. For the expected available capacity to equal 1, existing capacity should exceed 1.

Firms' profits remain as in the basic model, with one caveat only: in case of a capacity deficit, firms make profits with probability $(1 - \gamma)$. This reduces scarcity rents, which in turn leads to lower equilibrium capacity as compared to the basic model. In line with our previous analysis, the market generates under-investment with respect to the first-best, and the distortion is increasing in γ. Removing the price cap by setting $P = v$ would alleviate but not fully offset the inefficiency. Thus, even if the assumptions of the energy-only market paradigm were satisfied (free entry and no market power), it would be inefficient to rely on scarcity pricing as a way to promote investments (Llobet and Padilla (2018) make the same point).

The following Proposition summarizes the results of this section.

Proposition 1 *In equilibrium, aggregate capacity is*

$$K^* = \frac{2}{2 - 3\gamma} \frac{P(1 - \gamma) - c}{P} < K^{FB}.$$

An increase in γ reduces K^* and enlarges the market distortion with respect to the first-best, $K^{FB} - K^*$. Setting $P = v$ does not close the gap.

2.4 Market power

In contrast to the energy-only market paradigm, the free entry and the no market power assumptions are rarely satisfied in practice, as the empirical literature and the case law have broadly documented. In this section we relax these two assumptions in order to characterize equilibrium investments (for simplicity, we again rule out blackouts).

To capture the notion of market power in the energy market, we assume that firms $i = 2, ., n$ bid at marginal cost while firm 1 bids strategically, i.e., so as to maximize its profits over the residual demand given its rivals' supply.[12] The latter will be referred to as the *dominant firm*, while the former will be referred to as the *fringe firms*. We will use k_F to denote the aggregate capacity of the fringe firms, i.e., $k_F = \sum_{f=2}^{N} k_f$, and $K = k_1 + k_F$ to denote the aggregate capacity of all firms in the market. All firms choose their capacities strategically.

In order to characterize the Subgame Perfect Nash Equilibria, we proceed by backwards induction: first, we characterize equilibrium pricing for given capacities, and subsequently we characterize equilibrium investments.

Pricing stage Since the fringe firms bid at marginal cost, equilibrium prices equal zero whenever the fringe capacity is enough to satisfy demand (if $\theta \leq k_F$). Otherwise, the dominant firm is able to raise the market price up to the price-cap. The dominant firm either serves the residual demand if $k_F \leq \theta \leq K$, or it sells up to capacity if there is not enough capacity overall (if $\theta \geq K$). Hence,

---

[12]Allowing all firms to price strategically would result in equilibria with the same feature: in all equilibria, all firms but one produce the same as if they bid at marginal cost, whereas the remaining firm bids so as to maximize its profits over the residual demand (see Fabra *et al.* (2006) and Fabra *et al.* (2011)). Exogenously fixing the identity of the firm that (possibly) bids above marginal cost allows to simplify the model without altering the qualitative nature of the results. Appendix B contains a detailed analysis of the case in which all firms behave strategically.

profit expressions have three terms that we can label respectively as (i) market power rents, (ii) scarcity rents (as defined above) and (iii) investment costs.[13] Formally, expected equilibrium profits for the dominant firm and for the fringe firms are given by

$$\pi_1 = P \int_{k_F}^{k_F+k_1} [\theta - k_F] \, d\theta + P \int_{k_F+k_1}^{1} k_1 d\theta - ck_1 \tag{4}$$

$$\pi_f = P \int_{k_F}^{k_F+k_1} k_f d\theta + P \int_{k_F+k_1}^{1} k_f d\theta - ck_f, \text{ for } f = 2,.,n. \tag{5}$$

Investment stage Let us first consider the dominant firm's marginal incentives to invest. Taking derivatives in equation (4),

$$\frac{\partial \pi_1}{\partial k_1} = P[1 - K] - c, \tag{6}$$

shows that the dominant firm only benefits from capacity expansions in scarcity periods, i.e., when $\theta > K$, and event which occurs with probability $1 - K$. In all other instances, the firm is not capacity constrained, and hence increases in its capacity would not lead to greater production. In turn, since this implies that the dominant firm's incentives to invest only depend on total investment, condition (4) fully determines equilibrium aggregate capacity:

$$K^* = \frac{P - c}{P} \leq K^{FB}. \tag{7}$$

The above expression again shows that unless price caps are removed ($P = v$), there is underinvestment relative to the first best. Interestingly, since aggregate capacity is increasing in P, there is a trade-off between mitigating market power (which calls for lower price caps) versus inducing investment incentives closer to the first-best (which calls for higher price caps).

Let us now turn to the fringe firms' incentives to invest. Differentiating equation (5) with respect to k_f, taking as given the capacities of all the other firms,

$$\frac{\partial \pi_f}{\partial k_f} = Pk_1 - Pk_f + P[1 - K] - c. \tag{8}$$

The fringe firms benefit more from expanding their capacity as compared to the dominant firm, given that they tend to produce at capacity more often (first term in condition (8)). However, an expansion in the fringe firms' capacity also increases the probability of marginal cost pricing, which tends to discourage investment (second term in (8)).

Imposing symmetry among the fringe firms, equilibrium capacities are given by

$$k_f^* = \frac{1}{n} \frac{P - c}{P} \text{ for } f = 2, ..., n. \tag{9}$$

---

[13]Importantly, we will see that at equilibrium capacities, investment costs are covered by scarcity rents, implying that equilibrium profits are equal to market power rents only.

Since at $k_1 = k_f$, the first-order condition of fringe firms (expression (8)), bolts down to that of the dominant firm (expression (8)), it follows that in equilibrium, all firms choose symmetric capacities.[14] Hence, $k_i^* = k^*$ for all $i = 1, ..., n$.

The following Proposition summarizes the results so far.

Proposition 2 *In equilibrium, each firm's capacity equals $k^* = \frac{1}{n}\frac{P-c}{P}$. Aggregate capacity is below the first-best level, unless $P = v$.*

Using the expressions for equilibrium capacities, we can compute firms' equilibrium profits. In particular, since firms invest up to the level at which scarcity rents just cover investment costs, profits are simply given by market power rents (i.e., the second and third terms in expressions (4) and (5) cancel out):[15]

$$\pi_1^* = P \int_{(n-1)k^*}^{nk^*} [\theta - (n-1)k^*] d\theta = \frac{1}{2}\frac{P}{n^2}\left(\frac{P-c}{P}\right)^2 \tag{10}$$

$$\pi_f^* = \int_{(n-1)k^*}^{nk^*} k^* d\theta = \frac{P}{n^2}\left(\frac{P-c}{P}\right)^2. \tag{11}$$

Firms' profits are always positive (although they tend to zero as the number of firms becomes very large). Fringe firms make more profits than the dominant firm given that they produce at capacity more often.

In turn, since total capacity is the same as under the energy-only paradigm, so is total welfare.[16] However, since firms make market power rents, consumers surplus is lower by the same amount. This leads to an important conclusion. In contrast to the energy-only paradigm, consumer surplus is not maximal at $P = v$: whereas this would align investment incentives with the first-best, prices would be too high from consumers' point of view. It is preferable to set $P < v$ even if it implies that capacity is distorted downwards.[17] Hence, in the presence of market power, it is not optimal to remove price caps. Market power - rather than price caps per se - imply that at the second best solution, equilibrium aggregate investment is inefficiently low.

---

[14]The equilibrium is symmetric under uniformly distributed demand, but not necessarily so under a general demand distribution G. In this case, the equilibrium aggregate capacity is $G^{-1}\left(\frac{P-c}{P}\right)$ and market structure is be determined by the following FOC

$$G\left((n-1)k_f\right) + k_f G'\left((n-1)k_f\right) = \frac{P-c}{P}.$$

Furthermore, the equilibrium also involves asymmetric capacities (even with uniformly distributed demand) when all firms are assumed to behave strategically. See Appendix B.

[15]When comparing outcomes with those under the free entry condition (derived in the previous section), note that the number of firms is not the same in the two analysis (in that section it is endogenous because of the free entry condition, in this one it is taken as given). Also note that, in terms of profits, profits are now higher because firms make positive profits also when demand is above the capacity of the fringe but below total capacity.

[16]If demand was price-elastic, market power would generate a deadweight loss. Hence, removing price caps would not allow to maximize welfare.

[17]To see this, it suffices to check that the first derivative of CS with respect to P, evaluated at v, is negative. This implies that CS increases when P is reduced below v. See also Fabra *et al.* (2011).

Proposition 3 *To maximize consumer surplus, it is optimal to set $P < v$. The resulting aggregate capacity is below the first-best level.*

3   Capacity Mechanisms

We have just shown that market power erodes consumer surplus: either investment incentives are optimal but prices are too high, or prices are closer to marginal costs but there is under-investment. A single instrument (price caps) is simply unable to serve the double purpose of inducing the right investment incentives while avoiding market power concerns. How can the regulator reconcile these two objectives?

In this section we show that capacity payments (i.e., payments that are a function of capacity regardless of firms' production) can potentially play that role.[18] In particular, we will show that the combination of capacity payments and price caps allows to disentangle the incentives to invest (which are a function of capacity payments) from the market power concerns (which are a function of price caps).

Types of capacity mechanisms   We can distinguish two types of capacity mechanisms, depending on whether the regulator chooses prices or quantities: (i) capacity prices (price regulation): the regulator pays an extra price s per unit of capacity and, given that price, investors choose their capacities; or (ii) capacity markets (quantity regulation): the regulator decides how much capacity is needed, and runs auctions (so called, capacity markets) to determine the price s that investors require to build the new capacity.[19]

In a world in which the regulator has complete cost information and there is no market power in the capacity market,[20] both types of capacity instruments are equivalent to each-other (in line with Weitzman (1974)'s seminal contribution). As Cramton *et al.* (2013) put it, the choice between these two basic approaches, either price or quantity regulation, is not a choice between a market approach and a regulated approach. Rather, the choice between the two depends on other factors such as risk attitudes, market power, or the coordination of investments in capacity. In what follows, since we assume complete cost information and no market power in the capacity market (until otherwise stated), it is inconsequential whether we assume that capacity payments take the form of either price or quantity regulation.

---

[18]Our model assumes that there is no pre-existing capacity in the market. In the next section we analyse the case in which a capacity market is introduced in a market in which there is already some existing capacity.

[19]Within these family of instruments, regulators also resort to bundling capacity with financial instruments, as analyzed further below.

[20]Another reason why price and quantity regulation might not be equivalent is uncertainty about the true demand distribution. So far we have assumed that both the regulator as well as firms share the same belief about the demand distribution. Disparities in both beliefs break the equivalence between price and quantity regulation.

3.1 Equilibrium investment

Firms receive a capacity payment s per unit of capacity, which they take as given. Accordingly, to compute firms' profits, it suffices to add the term sk_i, for $i = 1, ..., n$, to the profit expressions (4) and (5). Since the capacity payment is equivalent to a capacity subsidy, the actual per unit cost of capacity now becomes $c' = c - s$.

Capacity revenues are already fixed by the time firms submit their bids to the wholesale market. Hence, while s does not have a direct impact on equilibrium pricing, it affects pricing through its impact on investment incentives. In particular, s increases aggregate capacity, which in turn reduces energy prices. Indeed, using the equilibrium expressions above,

$$K^* = \frac{P - c + s}{P}. \tag{12}$$

Under price regulation, if the regulator wants firms to choose capacity K^*, she has to set a capacity price that covers firms' investment costs net of the scarcity rents (firms keep the market power rents, as noted before). In other words, scarcity rents just cover the firm's implicit cost of capacity. From (12),

$$s = c - (1 - K^*) P. \tag{13}$$

The same solution would be achieved under quantity regulation if the regulator demands K^* in a capacity market. In this case, the market would also clear at (13).

If the regulator wants to induce firms to invest up to the first best capacity K^{FB}, then

$$s^{FB} = c \frac{v - P}{P},$$

which is decreasing in P: it involves capacity payments covering the full investment cost ($s = c$) when prices are capped at marginal cost ($P = 0$), or no capacity payments ($s = 0$) when price caps removed ($P = v$). The latter result is in line with the energy-only paradigm. However, as we will later show, this solution is sub-optimal for consumers.

An implementation challenge: incomplete cost information As an aside, to implement such solutions, the regulator needs to know investment costs c.[21] However, when the regulator faces asymmetric information, price and quantity regulation are no longer equivalent. For instance, under price regulation, if the true investment cost is c but the regulator believes it is $\widehat{c}$ and therefore offers to pay $s = \widehat{c} - (1 - K^*) P$, the market will deliver capacity

$$\frac{P - c + s}{P} = K^* + \frac{\widehat{c} - c}{P}$$

---

[21]If we had not normalized production costs to zero, it would be evident from the above expressions that the regulator would also need to know production costs.

i.e., too much capacity (above the regulator's desired K^*) if the true costs are below the regulator's expectation ($c < \widehat{c}$), or too little otherwise.

Alternatively, under quantity regulation, if the regulator procures K^*, the capacity market will clear at $s = c - (1 - K^*) P$, i.e., at a capacity price that is lower than expected if $\widehat{c} > c$, or higher otherwise. This might lead to inefficiencies to the extent that, had the regulator known that the true cost was lower, his desired capacity target would have been higher (we will later see that the optimal capacity depends on how much it costs to induce firms to invest). The use of downward sloping demand functions in the capacity market can partly mitigate this, with the additional benefit of mitigating market power.

3.2 How do capacity payments affect market outcomes?

Expected prices For given s, expected prices at equilibrium capacities are given by

$$
E\left[p\right] = P \int\limits_{(n-1)k^*}^{1} d\theta = c + \frac{P - c}{n} - \frac{n - 1}{n} s.
$$

Overall, expected energy prices are decreasing in s given that s increases total capacity and hence intensifies competition. In this sense, capacity payments mitigate a two-fold market failure: inefficiencies in capacity choices and in price setting.

Firms' profits Such a price depressing effect has a negative impact on firms' energy market profits. However, firms also benefit as they receive capacity payments and produce more (there is less demand rationing). What is the net effect of capacity payments on their profits? Again, replacing c by $c' = c - s$ in the profit expressions (10) and (11), allows to conclude that equilibrium firms' profits are increasing in s. Hence, firms benefit from capacity payments.

The above conclusion relies on the assumption that as s is introduced the price cap P remains unchanged. However, as we will see later, an increase in s allows to reduce P. Hence, the overall impact on firms' profits will depend on whether and by how much an increase in s allows for a reduction in P.

3.3 The optimal policy: price caps and capacity payments

Capacity payments create a trade-off for consumers: they are costly, but they depress market prices and allow for greater consumption. The following Proposition characterizes the optimal solution to this trade-off.

Proposition 4 *For given P, consumer surplus is maximized at*

$$
s^* = c - P \frac{(2n - 1) P + cn^2}{(2n - 1) P + vn^2}.
$$

Consumer surplus is concave in s because of two countervailing effects: an increase in s leads to greater consumption, but also to higher capacity payments. For low capacity prices (up to s^*), consumer surplus goes up as the increase in consumption dominates over the increase in payments. The opposite is true for high capacity prices (above s^*).[22] Hence, capacity payments do not necessarily make consumers worse off: they are worse off only when the capacity price is set too high.

The optimal capacity price s^* is decreasing in P, thus reflecting the trade-off between market power and investment incentives. If prices are constrained to be equal to marginal costs, $P = 0$, the capacity price covers investment costs, $s^* = c$. If price caps are removed, $P = v$, s^* becomes negative: consumers are better off if the regulator asks firms a fee to enter the market, even if this comes at the cost of distorted investment incentives.

Given the trade-off between price caps and capacity payments, a natural question arises: what is the optimal combination between the two instruments? Since the iso-CS curves (i.e., the (P, s) pairs that result in the same level of consumer surplus) are non-linear, there is not a one to one correspondence between reducing the price cap and increasing capacity payments. The following Proposition characterizes the optimal (P, s) pair.

Proposition 5 *Consumer surplus is maximized at $P^* = 0$ and $s^* = c$.*

At the optimal capacity price, s^*, consumer surplus is decreasing in P. Hence, the optimal solution has $P^* = 0$ and $s^* = c$. Essentially, this involves capping energy prices at marginal costs and fully subsidizing investment costs. This solution results in too much capacity (equilibrium capacity would equal 1, which exceeds first-best capacity), unless the regulator puts a cap on how much capacity she wants to pay at s. For instance, the first-best capacity can be implemented by setting $P = 0$ and paying $s = c$ only up to the first-best capacity. Firms cover all their energy costs with their energy market revenues, and their investment costs with their capacity payments.

If the regulator maximizes a weighted sum of consumer surplus and profits, a less stringent policy is optimal, i.e., let some market power be exercised by increasing the price cap above marginal costs, but do not fully subsidize the fixed cost of the investment.

Even if appealing, the above solutions are difficult to implement in practice as the regulator faces a fundamental problem: asymmetric information about firms' production and investment costs. The second-best solution under asymmetric information (e.g. if the regulator does not observe production costs) would involve a price-cap above marginal costs, allowing firms to make positive expected profits. Furthermore, in practice, there coexist various generation technologies, implying that a single price cap would not lead to zero profits for all - unless there are technology-specific price caps, an issue to which we will return below.

---

[22]Expected payments by consumers are always increasing in s.

4 Further Issues

4.1 Targeted versus market-wide mechanisms

So far, we have assumed no pre-existing capacity in the market. We now want to understand the effects of introducing capacity payments in a sector in which there already exists capacity K_t. A natural question then arises: should the regulator adopt targeted mechanisms (i.e., give support only to the new capacity) or market-wide mechanisms (i.e., give support to all market participants)? Examples of quantity-based targeted mechanisms are strategic reserves (i.e., some capacity is left outside to be dispatched by the System Operator in stress situations), or tenders for new capacity. Examples of quantity-based market-wide mechanisms are capacity markets[23] or decentralized obligation schemes that put the capacity obligation on the load-serving entities. Price-based capacity payments can be either targeted or market-based depending on whether only the new plants or all of the existing ones are entitled to receive them.[24] To address this question, since the market is assumed to be in equilibrium prior to introducing capacity payments, we let existing capacities be those in Proposition 2.

Importantly, the first point to note is that investment incentives depend on marginal profits, not on profit levels. Hence, equilibrium aggregate capacity is given by expression (12), regardless of whether capacity payments are paid to all plants or only to the new ones. In contrast, consumer surplus is higher when only new plants receive capacity payments simply because consumers pay less capacity payments to obtain the same consumption and energy price levels. In turn, this allows the regulator to optimally choose a higher s when only new plants receive capacity payments. Ultimately, this results in capacity choices closer to the first-best.

Proposition 6 *Let s^*_{NEW} be the optimal capacity payment when only new plants receive capacity payments. Then*

$$s^*_{NEW} - s^* = \frac{P - c}{\frac{2n-1}{n^2} + \frac{v}{P}} > 0.$$

Hence, aggregate equilibrium investment is closer to K^{FB} when only new plants receive capacity payments,

$$K^*_{NEW} = \frac{P - c + s^*_{NEW}}{P} > \frac{P - c + s^*}{P} = K^*,$$

and consumers are better off.

Can firms claim that capacity payments hurt them if only new plants receive them? Firms' profits with capacity payments for all plants are given in equations (10) and (11). As we showed before, these profits increase in s. If only new plants receive s, profits are diminished by the

---

[23]Some capacity market designs are market-wide in the sense that all plants can participate in them. However, the commitments for old and new plants differ as the former are typically entitled to receive capacity payments for shorter periods of time (e.g., in the UK market, 1 year for the existing plants and 15 years for the new ones).

[24]See Section 5 of the EC's sector inquiry, (European Commission, 2016).

payments that old plants would have received, sK_t. In particular,

$$
\begin{aligned}
\pi_1^T &= \frac{1}{2n^2}\frac{(P-c+s)^2}{P} - \frac{s}{n}\frac{P-c}{P} \\
\pi_f^T &= \frac{1}{n^2}\frac{(P-c+s)^2}{P} - \frac{s}{n}\frac{P-c}{P}.
\end{aligned}
$$

These profit expressions are now decreasing in s given that a higher s induces more capacity investments and hence lower prices. Therefore, the change in profits before and after the introduction of the capacity payments when only new plants receive them are given by

$$
\begin{aligned}
\pi_1 - \pi_1^T &= \frac{s}{n}\frac{P-c}{P} - \frac{s}{2n^2}\frac{2(P-c)+s}{P} < sk \\
\pi_f - \pi_f^T &= \frac{s}{n}\frac{P-c}{P} - \frac{s}{n^2}\frac{2(P-c)+s}{P} < sk.
\end{aligned}
$$

The amount of profits that firms lose is less than the capacity payments they would have received under a market-wide capacity payment. In other words, paying for the old plant's capacity not only leaves investment unchanged, but it also implies an overcompensation beyond the profits firms lose when capacity payments are introduced.

If one focuses on the profits of the old plants only (rather than on the profits of firms' portfolio) certainly the old plants lose because they do not receive s and sell their output at lower expected prices. However, this profit loss is partly compensated by the (strictly) positive profits of the new capacity.

4.2   Market power in the capacity market

So far we have assumed that there is no market power in the capacity market, i.e., firms take s as given when taking their capacity choices. However, this assumption need not hold in practice: just as firms exercise market power in the energy market, there is empirical evidence of the exercise of market power in capacity markets.[25]

To capture this, let us now assume that the fringe takes s as given, but the dominant firm does not. Hence, the supply by the fringe firms is the same as above, and the market clearing condition when the demand in the capacity market is vertical at K, is given by

$$
K = \frac{n-1}{n}\frac{P-c+s}{P} + k_1. \tag{14}
$$

Solving for s gives the capacity price that clears the market:

$$
s = c - (1-K)P + P\frac{n}{n-1}\left(\frac{K}{n} - k_1\right).
$$

---

[25]Schwenen (2015) analyzes the New York capacity market and Teirilä (2017) analyzes the auctions for reliability options in Ireland.

The elasticity of the capacity choice by the fringe (the higher s, the higher its capacity supply) introduces demand elasticity in the residual demand faced by the dominant firm, even if the market demand for capacity is vertical. The dominant firm thus maximizes profits, given the fringe's supply and the demand at the capacity market. The resulting capacity payment is the same expression as in expression (13), with the addition of the third term, which is strictly positive whenever the dominant firm withholds part of its capacity.

The following proposition indeed shows that capacity withholding is a profit-maximizing strategy for the dominant firm.

Proposition 7 *In equilibrium, a capacity market with vertical demand K^* clears at capacity choices*

$$k_1^* = \frac{K^*}{n+1} \text{ and } k_f^* = \frac{n}{n^2-1}K^*$$

and a capacity price

$$s^* = c - (1 - K^*)P + \frac{K^*}{n^2-1}P. \tag{15}$$

In equilibrium, the dominant firm withholds investment so as to drive the capacity price up as compared to when s is taken as given - comparing expressions (13) and (15), the capacity price is increased by the second term in the expression (15). As a consequence, the dominant firm becomes relatively smaller, since it now possesses a $(1/n+1)$th of total capacity instead of $(1/n)$th. This has a positive effect on consumers (given that the dominant firm is relatively smaller, energy prices go down), but also a negative one (capacity payments increase). Since the latter effect dominates, the overall effect of market power in the capacity market is to make consumers worse off.

Furthermore, the presence of market power in the capacity market implies that the optimal K^* goes down, i.e., market power in the capacity market reduces how much capacity the regulator chooses to procure. Indeed, comparing outcomes with and without market power in the capacity market at the aggregate capacity level that maximizes consumer surplus, shows that: (i) the optimal capacity is higher without market power in the capacity market; (ii) the equilibrium capacity price is higher without market power in the capacity market; and as a result, (iii) market power in the capacity market makes consumers worse off.

This suggests that consumers might be better off if the regulator submits a downward sloping demand for capacity which mitigates the dominant firm's incentives to mitigate market power in the capacity market. This is the case in the UK capacity market, where the demand schedule for capacity is downward-sloping, passing through the desired amount at the so-called net Cost of New Entry (CoNE) (Pollitt and Haney, 2013). This design is similar to that in the PJM (Bowring, 2013).

4.3   Reliability options

In some countries, regulators are increasingly bundling capacity payments with some sort of financial commitments. The objective is three-fold: provide investment incentives, mitigate market power, and incentivize the plants' availability.

The most commonly used contracts of this type are the so-called reliability options. These give investors a capacity payment (the option price) in exchange of their commitment to pay back any positive difference between the wholesale market price and the contract's strike price times the committed quantity. Reliability options thus provide investors with a certain flow of revenues, while consumers benefit from the commitment that prices will not be increased above the strike price. Typically, these contracts are allocated through auction mechanisms: bidders compete for the right to deliver energy under a reliability option with a given strike price, thus allowing the capacity price to be set competitively.

To formalize the effects of reliability options, let the contracted quantity equal the plant's capacity k, let f denote the contract's strike price and, just as before, let s represent the capacity price (or option price) that is set through the auction. What is the impact of reliability options on firms' bidding behavior?

A firm subject to a reliability option has to pay back the difference between the market price p and the strike price f whenever positive. Hence, profits become

$$\pi = pq - \max\{p - f, 0\} k - (c - s) k$$

or, equivalently,

$$\pi = \begin{cases} pq - (c - s) k & \text{if} \quad p \le f \\ fk - (c - s) k - p (k - q) & \text{if} \quad p \ge f \end{cases}$$

When the market price is below f, firms' bidding incentives remain unchanged. However, when the market price is above f, the firm faces a two-fold incentive. First, regarding pricing incentives, the reliability option puts the firm into a net-buyer position: it is *as if* the firm had to buy an amount of energy equal to its own capacity k but only sells $q < k$. As a net-buyer, the firm does not have incentives to bid above f. And second, regarding availability incentives, the firm is highly encouraged to produce up to capacity as failure to do so would imply an endogenous penalty $-p(k - q)$. As this penalty is harsher the higher the market price, the incentives for being available are greater during scarcity times.

Reliability options are auctioned off for a price s, which is set competitively. Hence, s compensates investors for the capacity costs not covered through the market. Furthermore, since the firm subject to the reliability contract does not want to raise the market price above f, the strike price acts as a plant-specific price cap. We thus obtain a similar expression for s as in equation (13) above, where f replaces P,

$$s^* = c - (1 - K^*) f.$$

The choice of f does not affect how much is invested, but it has an impact on the energy market price through the effects on market power. Just as our discussion regarding the choice of s and P demonstrated, a reduction in f makes the reliability option more costly but it saves consumers more than it costs, as the reliability option becomes more valuable in preventing market power.

Thus, in line with Proposition 5, the optimal reliability option has a strike price equal to marginal costs, $f = 0$. Furthermore, if there is enough competition in the auction for reliability options, s will be driven down to investment costs, $s = c$.

If not all firms are subject to reliability options, some market power will be exercised in the energy market. Since this will imply that p will rise above f with some probability, the resulting s will be lower than c as the firm will be able to recover part of its investment costs through the energy market. Thus, even if some market power is exercised, the market power rents of the firms subject to reliability options will be competed away through the auction.

In this simple model we have abstracted from risk issues. However, to the extent that investors are risk averse in practice, an additional benefit of bundling physical investment with financial commitments is the reduction in risk premia. First, firms receive a constant and certain payment for their capacity, as opposed to receiving uncertain market revenues in periods of scarcity. And second, under reliability options, the energy price received by firms is volatile only when the market price is between the firm's marginal costs and the option's strike price. Hence, a strike price close to marginal costs also reduces the plant's risk exposure, contributing to the reduction in risk premia.

5   Conclusions

The energy-only market paradigm relies on two key assumptions that are typically not satisfied in practice: free entry and no market power in the energy market. Once these are relaxed, the conclusion is unambiguous: relying on scarcity pricing as a way to promote investments is not efficient. While removing price caps allows for optimal capacity choices, this comes at the cost of market power in the energy market. If with some probability involuntary demand rationing (i.e., system blackouts) cannot be avoided, or if energy demand is downward sloping, energy-only markets do not achieve the first-best, even if price caps are removed.

The energy-only market paradigm is right in pointing out that price caps in the energy market lead to under-investment with respect to the first-best. However, it is important to stress that such inefficiency is not created by price caps. Rather, inefficiencies are created by market power, which price caps are meant to mitigate. Maximization of consumer surplus calls for binding price caps (i.e. below consumers maximum willingness to pay) even if these lead to under-investment and thus lower consumption. If energy demand depicts some elasticity, price caps also allow to foster consumption and thus increase welfare.

The trade-off between providing the right investment incentives and mitigating market power cannot be disentangled with one regulatory instrument only: if price caps are set close to consumers' maximum willingness to pay so as to provide the correct investment incentives, the resulting market prices are too high; while if price caps are set close to marginal costs so as to mitigate market power, the resulting investments incentives are too weak. Disentangling both objectives thus calls for combining price caps with additional instruments, such as capacity payments: the latter contribute to covering investment costs net of energy market profits, while the former allow to mitigate market

20

power. Ultimately, this leads to more efficient capacity choices and greater consumer surplus. Importantly, capacity payments do not have a direct impact on energy prices. They impact energy prices only through their impact on capacity: capacity payments promote more investment in capacity, which in turn promotes competition and hence drives energy prices down.

The efficiency improvements triggered by capacity payments might be hampered in the presence of market power in the capacity market. Market power raises the costs of procuring a given amount of capacity, which might in turn induce the regulator to optimally procure less capacity than in the absence of market power. Hence, even when capacity payments are a pure transfer between consumers and producers, market power in the capacity market may end up having a negative effect on total welfare through the choice of the regulator's optimal capacity target. Using downward sloping demand functions in the capacity market might alleviate market power concerns.

Last, this paper has demonstrated the advantages of bundling capacity payments with financial commitments, such as reliability options. Not only these provide a certain flow of revenues to the investors - a key issue to support investment in capital-intensive long-lived assets - but also help mitigate market power in the energy market. Importantly, this issue is closely linked to the parameters chosen by the regulator. The closer is the strike price to marginal costs, the more valuable reliability options become in preventing market power.

Increasing demand response can certainly play a key role in facilitating security of supply at least cost. If consumers are faced with real time prices, they might be encouraged to shift their load from peak to off-peak periods, thus reducing the need to maintain excess capacity while at the same time mitigating market power concerns. In this sense, promoting demand responsiveness should be viewed as complementary to capacity support mechanisms- the extent of this complementarity is yet to be empirically demonstrated.

References

[1] Bajo-Buenestado, R. (2017) Welfare Implications of Capacity Payments in a Price-Capped Electricity Sector: a case study of the Texas market (ERCOT), *Energy Economics*, 64 (C), 272-285.

[2] Bushnell, J., Flagg, M. and E. Mansur (2017), Capacity Markets at a Crossroad, Energy Institute at Hass Working Paper WP 278.

[3] Borenstein, S. (2000) Understanding Competitive Pricing and Market Power in Wholesale Electricity Markets, The Electricity Journal 13(6), 49-57.

[4] Bowring, J., (2013), Capacity Markets in PJM, Economics of Energy and Environmental Policy 2(2), 47-64.

[5] Crampes, C. and D. Salant (2018) A Multi-Regional Model of Electric Resource Adequacy, Toulouse School of Economics WP 17-877.

[6] Cramton, P., Ockenfels, A. and S. Stoft (2013) Capacity Market Fundamentals, *Economics of Energy & Environmental Policy* 2(2).

[7] European Commission (2016), Final Report of the Sector Inquiry on Capacity Mechanisms, November 2016, COM(2016) 752 final.

[8] Fabra, N. and A. Creti (2007), Supply Security and Short-Run Capacity Markets for Electricity, *Energy Economics* 29(2), 259-276.

[9] Fabra, N., von der Fehr, N., and de Frutos, M. (2011) Market Design and Investment Incentives, *Economic Journal* 121, 1340–1360.

[10] Fabra, N., von der Fehr, N., and Harbord, D. (2006) Designing Electricity Auctions, *Rand Journal of Economics*.

[11] Hogan, W. (2005) On an "Energy only" Electricity Market Design for Resource Adequacy, California ISO.

[12] Joskow, P. L. (2007), Competitive Electricity Markets and Investment in New Generating Capacity, D. Helm (Ed.), The New Energy Paradigm, Oxford University Press.

[13] Joskow, P. L. and J. Tirole (2007), Reliability and Competitive Electricity Markets, *Rand Journal of Economics*, 38 (1), 60–84.

[14] Lambin, X. (2016), The Social Value of Demand Response and its Integration in Capacity Mechanisms, Working Paper, TSE, November.

[15] Lambin X. and T.-O. Leautier (2018), Capacity Mechanisms across Borders: who is Free Riding?, Working Paper, TSE, June.

[16] Llobet, G. and J. Padilla (2018) Conventional Power Plants in Liberalized Electricity Markets with Renewable Entry, forthcoming, *The Energy Journal.*

[17] Neuhoff, K., J. Diekmann, F. Kunz, S. Rüster, W.-P. Schill and S. Schwenen (2016) A Coordinated Strategic Reserve to Safeguard the European Energy Transition, *Utilities Policy* 41, 252-263.

[18] Newbery, D. (2016) Missing Money and Missing Markets: Reliability, Capacity Auctions and Interconnectors, *Energy Policy*, 401-410.

[19] Oren, S. (2005) Ensuring Generation Adequacy in Competitive Electricity Markets, in Electricity Deregulation: Choices and Challenges, 388-414.

[20] Pollitt, M.G., and A. Brophy Haney (2013) Dismantling a Competitive Electricity Sector: The U.K.'s Electricity Market Reform, The Electricity Journal 26(10).

[21] Schwenen, S. (2015) Strategic bidding in multi-unit auctions with capacity constrained bidders: the New York capacity market, Rand Journal of Economics, 46(4), 730–750.

[22] Teirilä, J. (2017) Market Power in the Capacity Market? The Case of Ireland, Cambridge Working Paper in Economics, 1727

[23] Weitzman, M.L. (1974), Prices versus Quantities, *Review of Economic Studies*, 41(4), 477-491.

[24] Zottl, G. (2011) On Optimal Scarcity Prices, *International Journal of Industrial Organization*, 29(5), 589-605.

Appendix A: Adding renewables

Within the basic model described in this paper, one can introduce renewables. Renewable production is denoted r and it is uniformly distributed in the interval $[0, R]$, where R is total renewable capacity. Given this assumption, average renewable availability is $R/2$. For simplicity, we assume no correlation between demand and renewable production.

Total Welfare can be expressed as:

$$W = \frac{v}{R} \int_0^R \left(\int_0^{K+r} \theta d\theta + \int_{K+r}^1 (K+r) \, d\theta \right) dr - cK - c_R R.$$

Taking renewable capacity R as given, the solution for K is:

$$K^{FB} = \frac{v-c}{v} - \frac{R}{2}.$$

Thus, if there is no renewable capacity ($R = 0$) we obtain the same solution as in the main model, equation (2). The higher R, the lower the optimal thermal capacity. Clearly, the more renewables there are, the less thermal capacity is needed. The extent to which R replaces K depends on several assumptions. Notably: the shape of demand distribution, the correlation between renewables production and demand, and whether blackouts are costly (see below). Under the simple assumptions of this note, the relationship is $1/2$ units of renewable capacity R are equivalent to one unit of thermal capacity K because the average availability of renewables is 50%. Even when $R = 1$ (so that total renewable capacity could cover the peak of demand if available), $K^{FB} > 0$ given the volatility of renewables.

To characterize equilibrium investment, let us suppose that the dominant firm does not posses any renewable capacity. Profits are:

$$\pi_1 = \frac{P}{R} \int_0^R \left(\int_{k_F+r}^{k_F+k_1+r} (\theta - r - k_F)\, d\theta + \int_{k_F+k_1+r}^1 k_1 d\theta \right) dr - ck_1$$

$$\pi_F = \frac{P}{R} \int_0^R \left(\int_{k_F+r}^1 k_F d\theta \right) dr - ck_F$$

Taking derivatives, we can characterize equilibrium aggregate investment and market structure:

$$K^* = \frac{P-c}{P} - \frac{R}{2}$$

$$k_1^* = k_i^* = \frac{1}{n}\left(\frac{P-c}{P} - \frac{R}{2} \right) \text{ for } i = 1, ..., n$$

Again, the first-order condition of the dominant firm determines aggregate capacity, which is the same as in the main model, Proposition 1, with the only caveat that the average availability of renewables is netted out from the optimal investment. The difference with the first-best capacity remains the same, given that the $R/2$ term appears in the two expressions. All our previous conclusions regarding the impact of price caps, the need for capacity payments and so on, thus apply equally here.

Appendix B: Symmetric strategic firms

In the basic model, we assumed that all but one firm bid at marginal costs (fringe firms) while the remaining firm sets the price that maximizes its profits over the residual demand (dominant firm). Instead, we now assume that all firms bid strategically. For simplicity, we let $n = 2$.

Let k^- and k^+ denote the capacities of the small and large firms in the market, respectively. As shown in Fabra *et al.* (20016), the pricing equilibrium can be characterized as follows: (i) if $\theta \leq k^-$, both firms bid at marginal cost; (ii) if $k^- < \theta \leq k^+$, the large firm bids at P while the small

firm bids sufficiently low to make undercutting by the large firm unprofitable; (iii) if $k^+ < \theta \leq K$, one of the two firms (either the small or the large) bids at P while the other one bids sufficiently low; (iv) in all equilibria both firms sell at capacity at P. Note that in case (iii), there are two equilibria, depending on which firm sets the market price. To make firms ex-ante symmetric, we assume that each of the two equilibria is played with equal probability. Thus, expected profits can be written as

$$\pi^- = P \int_{k^-}^{k^+} k^- \, d\theta + P \int_{k^+}^{K} \left[\tfrac{1}{2} k^- + \tfrac{1}{2} \left(\theta - k^+ \right) \right] d\theta + P \int_{K}^{1} k^- \, d\theta - ck^-,$$

$$\pi^+ = P \int_{k^-}^{k^+} \left(\theta - k^- \right) d\theta + P \int_{k^+}^{K} \left[\tfrac{1}{2} k^+ + \tfrac{1}{2} \left(\theta - k^- \right) \right] d\theta + \int_{K}^{1} k^+ \, d\theta - ck^+.$$

Since $\pi^- = \pi^+$ at symmetric capacity pairs, the expected profit function is everywhere continuous. Marginal returns to investment differ for large and small firms, thus creating a kink in firms' profit functions; in particular, the partial derivative of the profit function of a firm with respect to its own capacity 'jumps up' at the point where capacities are identical. As a result, best-replies do not cross the diagonal, implying that there cannot exist a symmetric equilibrium in capacity choices. Thus, the equilibrium must involve asymmetric capacity choices.

Solving the first order conditions, and checking that the second order conditions are satisfied, shows that in equilibrium

$$K^* = \frac{P - c}{P}$$

with

$$k^+ = \frac{3}{5} K^* > k^- = \frac{2}{5} K^*.$$

Thus, aggregate capacity is the same as in the basic model even though, unlike the basic model, the equilibrium involves asymmetric capacities.

If we add capacity payments s, the increase in aggregate capacity would mimic that of the basic model, while the degree of capacity asymmetry would remain as above. For given s, expected prices at equilibrium capacities would now be higher due to capacity asymmetries leading to more market power,

$$E\left[p\right] = P \int_{k^-}^{1} d\theta = P \int_{\frac{2}{5}\frac{P-c}{P}}^{1} d\theta = c + 3\frac{P-c}{5} - \frac{2}{5}s > c + \frac{P-c}{2} - \frac{1}{2}s.$$

For given P and s, consumer surplus is given by

$$CS = v \left(\int_{0}^{K^*} \theta \, d\theta + \int_{K^*}^{1} K^* \, d\theta \right) - P \left(\int_{\frac{2}{5}K^*}^{K^*} \theta \, d\theta + \int_{K^*}^{1} K^* \, d\theta \right) - sK^*.$$

Taking derivatives with respect to s shows that the optimal s^* is smaller than the one computed

in the basic model,

$$s^* = c - P\frac{21P + 25c}{21P + 25v} < c - P\frac{3P + 4c}{3P + 4v}.$$

Intuitively, an increase in s is less effective in reducing equilibrium prices because of capacity asymmetries. In any event, the main conclusions remain as in the basic model. Namely, s^* is decreasing in P; if $P = 0$ then $s^* = c$; and if $P = v$, s^* becomes negative (in this case, firms would be charged even a higher fee to enter the market given that they enjoy higher market power rents). Last, the optimal (s^*, P^*) continues to be $(c, 0)$.

We conclude that the qualitative conclusions of the dominant-fringe firms model and the model with symmetric strategic firms are the same.

Appendix C: Proofs of Propositions

Proof of Proposition 1:

Firms' profits are (note these are the same expressions as (4) and (5) above, with the single difference that $(1 - \gamma)$ now multiplies the second term):

$$\pi_1 = P \int_{k_F}^{k_F + k_1} [\theta - k_F] \, d\theta + (1 - \gamma) P \int_{k_F + k_1}^{1} k_1 d\theta - ck_1$$

$$\pi_f = P \int_{k_F}^{k_F + k_1} k_f d\theta + (1 - \gamma) P \int_{k_F + k_1}^{1} k_f d\theta - ck_f.$$

Taking the FOC of firm 1 implies that aggregate capacity is

$$K^* = \frac{2}{2 - 3\gamma} \frac{P(1 - \gamma) - c}{P},$$

which is evenly divided among all firms, $k_i^* = K/n$ for $i = 1, ..., n$.

The difference between the first-best capacity and the equilibrium capacity,

$$K^{FB} - K^* = \frac{1}{1 - 2\gamma} \frac{v(1 - \gamma) - c}{v} - \frac{2}{2 - 3\gamma} \frac{P(1 - \gamma) - c}{P} > 0$$

is increasing in γ.

Proof of Proposition 4: For given P and s, consumer surplus is given by

$$CS = v \left(\int_{0}^{nk} \theta d\theta + \int_{nk}^{1} nk d\theta \right) - P \left(\int_{(n-1)k}^{nk} \theta d\theta + \int_{nk}^{1} nk d\theta \right) - snk.$$

Plugging the equilibrium capacities k^* and taking derivatives:

$$\frac{\partial CS}{\partial s} = -\frac{1}{P}\left((P-c+s)\frac{2n-1}{n^2}+c-v\frac{c-s}{P}\right)$$

$$\frac{\partial^2 C}{\partial s^2} = -\frac{1}{P}\left(\frac{2n-1}{n^2}+\frac{v}{P}\right)<0.$$

Since CS is concave, the FOC gives the optimal subsidy given P :

$$s^* = c - P\frac{(2n-1)P+cn^2}{(2n-1)P+vn^2}.$$

Proof of Proposition 6:

Consumer surplus is given by

$$CS = v\left(\int_0^{\frac{P-c+s}{P}}\theta d\theta + \int_{\frac{P-c+s}{P}}^1 \frac{P-c+s}{P}d\theta\right) - P\left(\int_{\frac{n-1}{n}\frac{P-c+s}{P}}^{\frac{P-c+s}{P}}\theta d\theta + \int_{\frac{P-c+s}{P}}^1 \frac{P-c+s}{P}d\theta\right)$$

$$-s\left(\frac{P-c+s}{P}-x\frac{P-c}{P}\right).$$

where $x=1$ if only new capacity receives capacity payments, and $x=0$ if both old and new receive capacity payments.

Taking derivatives and equating to zero, one can find he optimal s as a function of whether only the new ($x=1$) or both new and old ($x=0$) receive capacity payments,

$$s(x) = \frac{(P-c)\left(x-2/n+1/n^2\right)+\frac{v-P}{P}c}{\frac{2n-1}{n^2}+\frac{v}{P}}.$$

The optimal $s(0)=s^*$ as before, and $s(1)=s^*_{NEW}$ where

$$s^*_{NEW}-s^* = \frac{P-c}{\frac{2n-1}{n^2}+\frac{v}{P}}>0.$$

In turn, the resulting capacity (for a given P) is higher and closer to the first-best,

$$K^*_{NEW} = \frac{P-c+s^*_{NEW}}{P} > \frac{P-c+s^*}{P} = K^*.$$

Proof of Proposition 7:

The dominant firm maximizes profits, given the fringe's supply $k_F = \frac{n-1}{n}\frac{P-c+s}{P}$,

$$\pi_1 = P \int_{\frac{n-1}{n}\frac{P-c+s}{P}}^{\frac{n-1}{n}\frac{P-c+s}{P}+k_1} \left(\theta - \frac{n-1}{n}\frac{P-c+s}{P}\right) d\theta + P \int_{\frac{n-1}{n}\frac{P-c+s}{P}+k_1}^{1} k_1 d\theta - ck_1 + sk_1$$

where

$$s = c - (1-K)P + P\frac{n}{n-1}\left(\frac{K}{n} - k\right).$$

with s defined as in equation (15). From the FOC, the solution is:

$$k_1^* = \frac{K}{n+1}$$

$$s = c - (1-K)P + \frac{KP}{n^2-1}$$

$$k_f^* = \frac{n}{n^2-1}K.$$

With market power in the capacity market, consumer surplus is:

$$CS = v\left(\int_0^K \theta d\theta + \int_K^1 K d\theta\right) - P\left(\int_{\frac{n}{n+1}K}^K \theta d\theta + \int_K^1 K d\theta\right)$$
$$- \left(c - (1-K)P + \frac{KP}{n^2-1}\right)K. \tag{16}$$

Without market power in the capacity market, consumer surplus is:

$$CS = v\left(\int_0^K \theta d\theta + \int_K^1 K d\theta\right) - P\left(\int_{\frac{n-1}{n}K}^K \theta d\theta + \int_K^1 K d\theta\right) - (c - (1-K)P)K.$$

Taking the difference between the two expressions,

$$P\int_{\frac{n-1}{n}K}^{\frac{n}{n+1}K} \theta d\theta - \frac{KP}{(n-1)(n+1)}K = -\frac{1}{2}K^2 P\frac{4n^2+n-1}{n^2(n-1)(n+1)^2} < 0,$$

shows that CS goes down when there is market power in the capacity market.

Taking derivatives in (16) (and checking that the SOC is satisfied), the optimal capacity to be procured for given P when there is market power in the capacity market is

$$K = \frac{(v-c)(n-1)(n+1)^2}{(P-v)(n+1) + 2Pn^2 + n^2v + n^3v}.$$

The resulting CS (evaluated at the optimal K when there is market power in the capacity market) becomes:

$$CS = \frac{1}{2} \frac{(v-c)^2 (n-1) (n+1)^2}{(P-v)(n+1) + 2Pn^2 + n^2 v + n^3 v}.$$

Similarly, taking derivatives (and checking that the SOC is satisfied), the optimal capacity to be procured for given P when there is no market power in the capacity market is

$$K = \frac{n^2 (v-c)}{-P + n^2 v + 2Pn}.$$

The resulting CS (evaluated at the optimal K when there is no market power in the capacity market) becomes:

$$CS = \frac{n^2}{2} \frac{(v-c)^2}{-P + n^2 v + 2Pn}.$$

Taking the difference between the two shows that consumers are worse-off under market power in the capacity market.