EPRG WORKING PAPER

UNIVERSITY OF CAMBRIDGE | Energy Policy Research Group

# Capacity mechanisms and the technology mix in competitive electricity markets

**EPRG Working Paper 1921**

**Cambridge Working Paper in Economics 1960**

## Par Holmberg and Robert A. Ritz

**Abstract** Capacity mechanisms are increasingly used in electricity market design around the world yet their role remains hotly debated. In this paper, we introduce a new benchmark model of a capacity mechanism in a competitive electricity market with many different generation technologies. We consider two policy instruments, a wholesale price cap and a capacity payment, and show which combinations of these instruments induce socially-optimal investment by the market. Changing the price cap or capacity payment affects investment only in peak generation plant, with no equilibrium impact on baseload or mid-merit plant. We obtain a rationale for a capacity mechanism based on the internalization of a system-cost externality – even where the price cap is set at the value of lost load. In extensions, we show how increasing renewables penetration enhances the need for a capacity mechanism, and outline an optimal design of a strategic reserve with a discriminatory capacity payment.

**Keywords** Investment, wholesale electricity market, capacity mechanism, capacity auction, strategic reserve

**JEL Classification** D41, L94

Contact Robert Ritz – rar36@cam.ac.uk

Publication July 2019

Financial Support Swedish Energy Agency, EPSRC (Grant number EP/R014604/1)

**www.eprg.group.cam.ac.uk**

# Capacity mechanisms and the technology mix in competitive electricity markets

Pär Holmberg* and Robert A. Ritz†,‡

This version: 29 June 2019

**Abstract**

Capacity mechanisms are increasingly used in electricity market design around the world yet their role remains hotly debated. In this paper, we introduce a new benchmark model of a capacity mechanism in a competitive electricity market with many different generation technologies. We consider two policy instruments, a wholesale price cap and a capacity payment, and show which combinations of these instruments induce socially-optimal investment by the market. Changing the price cap or capacity payment affects investment only in peak generation plant, with no equilibrium impact on baseload or mid-merit plant. We obtain a rationale for a capacity mechanism based on the internalization of a system-cost externality—even where the price cap is set at the value of lost load. In extensions, we show how increasing renewables penetration enhances the need for a capacity mechanism, and outline an optimal design of a strategic reserve with a discriminatory capacity payment.

*Keywords*: Investment, wholesale electricity market, capacity mechanism, capacity auction, strategic reserve

*JEL Classifications*: D41, L94

*Research Institute of Industrial Economics (IFN), Stockholm. Associate Researcher of the Energy Policy Research Group (EPRG), University of Cambridge, and Program on Energy and Sustainable Development (PESD), Stanford University. E-mail: par.holmberg@ifn.se

†Energy Policy Research Group (EPRG), Judge Business School, University of Cambridge. E-mail: r.ritz@jbs.cam.ac.uk

‡This work has been supported by the Swedish Energy Agency and the EPSRC (Grant number EP/R014604/1). We are grateful to audiences at the workshop on "Electricity systems of the future: Incentives, regulation and analysis for efficient investment" at the Isaac Newton Institute, University of Cambridge (March 2019) and the 8th Mannheim Energy Conference (May 2019). We are particularly grateful for comments from Eddie Anderson, Silvester van Koten and Andy Philpott. We thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the programme "Mathematics of Energy Systems" when work on this paper was undertaken.

# 1 Introduction

Capacity mechanisms are playing a growing role in electricity market design. In a nutshell, they award generators a capacity payment in exchange for being available to supply at a specified date. Capacity markets, in which this payment is determined by auction, are long-standing feature in several regional US power systems (such as ISO New England and PJM) and have in recent years been introduced in Great Britain, France, and Ireland. In 2018, the European Commission approved new market-wide capacity auctions in Italy and Poland as well as strategic reserves in Belgium and Germany.

There are a variety of justifications for a capacity mechanism. Its rationale is often said to arise from the presence of a price cap in the wholesale market. On one hand, a price cap protects electricity consumers from "too high" prices (perhaps resulting from the exercise of market power). On the other hand, setting it too low leads to underinvestment—known as the "missing money problem" (e.g., Joskow 2008). To this is added that greater renewables penetration reduces the running hours of conventional plant via the much-discussed "merit-order effect". A capacity mechanism, by providing generators with an additional revenue stream, has the potential to resolve the missing-money problem.

At the same time, the use and design of capacity mechanisms remains hotly debated. Some jurisdictions, such as Texas, rely on an "energy-only" market design without apparent need for capacity payments, some jurisdictions rely on a capacity auction and yet others use a strategic reserve to guarantee security of supply. Despite the recent proliferation of national capacity mechanisms, the European Commission has arguably taken a sceptical view due to concerns about market fragmentation and potential distortions of competition (EC 2016).[1] Moreover, some analysts speculate that the wholesale market will over time be eroded by zero marginal-cost renewables, with virtually all "action" shifting to the capacity market. In short, the debate around capacity mechanisms is one of the biggest policy issues for the future design of electricity markets.

In this paper, we introduce a new benchmark model of long-run investment with a capacity mechanism. Our main interest lies in understanding the optimal policy design when the regulator can use multiple instruments: a wholesale price cap and a capacity mechanism. We study three types of capacity mechanism: a capacity payment and capacity auction, both market-wide, and a targeted strategic reserve.

The key features of the model are as follows. First, we consider a wide range—technically, a continuum—of generation technologies, with the standard trade-off that a lower production cost comes with a higher investment cost.[2] This enables us to study how

[1] Capacity mechanisms also face a variety of practical challenges, ranging from the political influence on capacity procurement (Newbery & Grubb 2015), to the appropriate setting of de-rating factors for different generation technologies, to legal issues related to State Aid (EC 2016).

[2] Electricity markets are characterized by a range of generation technologies such as coal, diesel, hydro, natural gas, nuclear, oil, solar and wind. Within each technology, there are different types of plant in terms of size and efficiency. This means that, in practice, there could be a dozen or more individual

capacity mechanisms affect base-load, mid-merit and peak generation units in potentially different ways. Second, like much of the literature, we assume that demand is price-inelastic. This approximates real-world behaviour and makes the analysis tractable. We allow consumer demand be stochastic (which can also be interpreted as shifts in net demand due to variable renewable generation).[3] Third, if demand exceeds generation capacity, there is forced rationing, in the form of rolling black-outs, leading to a welfare loss for disconnected consumers. Moreover, we consider a system-cost externality which represents lost welfare due to accidental system-wide black-outs or that it is costly for the system operator to conduct rolling black-outs (Joskow & Tirole 2007; Fabra 2018; Llobet & Padilla 2018). Fourth, our interest lies in the optimal design of capacity mechanisms for the case of perfect competition among producers.[4]

We begin with the first-best benchmark for optimal investment. Social welfare consists of the gross consumer value from electricity minus production costs, investment costs and the system cost. A social planner keeps on investing until the marginal benefits of higher consumer value and a lower system cost are equal to the investment cost. A higher consumer value of lost load (VOLL) and a greater system-cost saving both lead to more investment into peaking plant.

We then study market-based investment under perfect competition. We show that there is a family of combinations of the price cap and capacity payment which achieves the social optimum via the market. One member of this family is setting the price cap at the VOLL and the capacity payment to internalize the system-cost externality. Establishing this policy family makes precise how much "uplift" in a capacity payment is needed to correct for different degrees of missing money.

In our model, there is a straightforward equivalence between (a) a capacity payment that leads to a market-based capacity volume or (b) a capacity procurement volume with a market-based capacity payment—akin to a capacity auction. A higher capacity payment and a higher price cap both raise investment and decreases the loss of load probability (LOLP); in this sense, these policy instruments are substitutes. Similarly, higher procured capacity leads to more investment and a higher capacity payment while, for a given procured capacity, a higher price cap reduces the "need" for a capacity payment.

A key observation is that these policy instruments work solely through their influence on peak plant. For baseload and mid-merit plant, the extra revenue from a higher capacity payment is exactly offset by the reduction in scarcity rent. The additional revenues, in equilibrium, go solely to financing new investment into peak plant.

We present two extensions to the benchmark model. First, we study how intermittent

"technology-types". We use a continuum to approximate this real-world diversity of discrete technologies.

[3]We do not attempt to model demand-side response (which is sometimes interpreted as a form of capacity mechanism) or electrical battery storage.

[4]Some of the finer details of capacity-market design are beyond our scope including the optimal setting of penalties for non-delivery and including reliability options (ROs) in the design.

renewables enhance the need for a capacity mechanism. Renewables crowd out conventional generation via a merit-order effect; all else equal, this exacerbates the system-cost externality by making it more difficult to control the power system. When "firm capacity" from conventional generation acts as a complement to intermittency, this raises the social value of investment in peaking plant—which is incentivized by a higher capacity payment.

Second, we outline a new socially-optimal design of a strategic reserve. A capacity payment that discriminates between plants inside and outside the reserve can easily lead to market distortions in investment. The key idea of our design is to avoid such inefficiencies by paying an extra-high price to non-reserve plants whenever the reserve is used.

**Contribution to the literature.** We contribute to a growing theoretical literature on capacity mechanisms. By considering a continuum of generation types, our approach differs from prior work which assumes a single representative technology (e.g. Léautier 2016; Fabra 2018) or two discrete technologies, sometimes interpreted as conventional and renewable generation (e.g. Llobet & Padilla 2018). Joskow & Tirole (2007) allow for a continuum of technologies but focus on cases with only a few demand outcomes—so producers invest only into 2-3 types of technologies.[5] Moreover, unlike us, previous literature on capacity markets often focuses on issues of market power.

Our work also relates to the classic literature on peak-load pricing which studies investment for a discrete set of technologies (see, e.g., Crew & Kleindorfer 1986). Our approach is similar to the model of Zöttl (2010) which, as far as we know, was first to use a continuous framework to study investment in electricity markets—but does not consider capacity payments or price caps.[6] For markets with inelastic demand, peak-load pricing with discrete technologies corresponds to the "screening curve analysis" which is widely used in the economics of electricity markets (Stoft 2002; Biggar & Hesamzadeh 2014; Léautier 2019). Our work simplifies existing results from this literature and broadens the analysis to include capacity mechanisms and system-cost externalities.

In sum, given these differences, we obtain several novel results including: (i) characterizing the family of socially-optimal combinations of a price cap and capacity payment, (ii) identifying the system-cost complementarity between renewables and conventional plant as the key driver of an enhanced need for a capacity mechanism, and (iii) deriving an equivalent socially-optimal design of a targeted strategic reserve.

**Plan for the paper.** Section 2 lays out our model. Section 3 characterizes the first-best outcome. Section 4 studies market-based investment with the policy instruments of a price cap and capacity mechanism. Section 5 analyzes the impacts of renewables penetration. Section 6 outlines a socially-optimal design of a strategic reserve. Section 7 concludes and discusses policy implications. Proofs are in the Appendix.

[5] Teirilä & Ritz (2019) present simulation results on the new Irish capacity market with several generation technologies. See Tangerås (2018) for an analysis of capacity markets in a multi-country setting.

[6] Zöttl (2010) also considers investments under Cournot oligopoly.

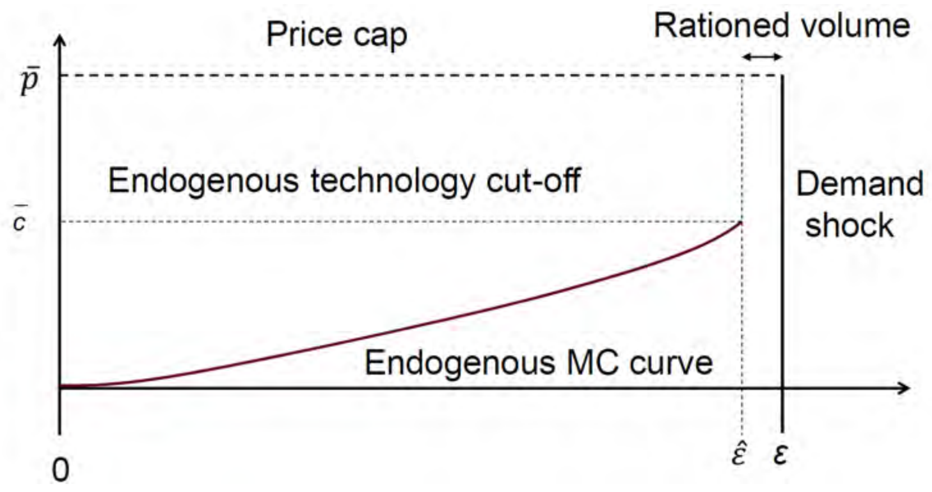


Figure 1: When demand $\varepsilon$ exceeds the market capacity $\hat{\varepsilon}$, there is demand rationing and the spot price is at the price cap $\bar{p}$.

## 2 The model

We consider a two-stage game with investments in Stage 1 and production in Stage 2. There is a continuum of production technologies with a continuum of marginal costs. We let $k(c)$ be the investment cost per unit of power for a technology with marginal cost $c \in [0, p^*]$, where $p^*$ is the reservation price (value of lost load, VOLL) for consumers. To get well-behaved solutions, we assume $k(c) \in (0, p^*)$, $k'(c) \in (-1, 0)$ and $k''(c) > 0$ for $c \in [0, p^*]$. Hence, as in practice, technologies with a higher marginal cost have a lower investment cost, $k'(c) < 0$, at a diminishing rate, $k''(c) > 0$. We also assume that consumers' VOLL is sufficiently high, $p^* > p_0^* = k(0)/[-k'(0)]$, which ensures that socially-optimal investments are positive. The total quantity invested in technologies with a marginal cost below $c$ is denoted by $q(c)$. The inverse of this supply function corresponds to a marginal cost curve $C'(q)$.

At the investment stage, there is uncertainty about consumers' electricity demand. In particular, demand $\varepsilon$ follows a probability distribution $F(\varepsilon)$ and density $f(\varepsilon)$ with the support $[0, \bar{\varepsilon}]$ where we assume $f(\varepsilon) > 0$ for $(0, \bar{\varepsilon})$. We can think of $F(\varepsilon)$ as being the fraction of time that demand is less than $\varepsilon$, in which case $F(\varepsilon)$ would be similar to a standard load duration curve.[7] We let $\bar{c}$ be the highest marginal cost for which there is investment. This technology cutoff is endogenously determined in our model. We let $\hat{\varepsilon} = q(\bar{c})$ be the corresponding total production capacity. As demand is assumed to be price-inelastic, forced demand rationing (rolling blackouts) is needed to keep the system in balance if $\varepsilon > \hat{\varepsilon}$ (that is, demand exceeds capacity). Hence, $1 - F(\hat{\varepsilon})$ represents the loss of load probability (LOLP). Figure 1 illustrates.

We assume that loss of load, in addition to the lost surplus of rationed consumers,

[7] The distribution $F(\varepsilon)$ can also be interpreted as *net* demand for conventional generation, taking into account production from renewables; we pursue this analysis further in Section 5.

has a system externality $M(\hat{\varepsilon})$. This might represent the expected cost of performing rolling black outs. Letting $J$ denote the realized system cost of conducting *controlled* rolling black-outs, which may be a function of the rationed volume $\varepsilon - \hat{\varepsilon}$, we have $M(\hat{\varepsilon}) = \int_{\hat{\varepsilon}}^{\overline{\varepsilon}} f(\varepsilon) J(\varepsilon - \hat{\varepsilon}) d\varepsilon$. An alternative interpretation of $M(\hat{\varepsilon})$ is as the (expected) welfare loss due to accidental *uncontrolled* system-wide black outs; more investment then improves system reliability—which is a public good. We assume that: (i) higher installed capacity reduces the system cost, at a decreasing rate, $M'(\hat{\varepsilon}) \leq 0 \leq M''(\hat{\varepsilon})$; and (ii) to avoid boundary solutions and yield an interesting analysis, the invested capacity $\hat{\varepsilon}$ must lie below $\overline{\varepsilon}$ (the highest demand realization), which will hold as long as $k(p^*) < [-M'(\hat{\varepsilon})]$.

# 3 Socially-optimal investment and the technology mix

We begin by solving the problem of a social planner who makes investment and production decisions in order to maximize social welfare. Social welfare is comprised of three components. First, it can be shown that the total investment cost in the first stage is given by:

$$K = \int_0^{\overline{c}} k(c) q'(c) dc. \tag{1}$$

We see that $q'(c)$ is essentially a density function. For small $\Delta c$, $q'(c) \Delta c$ is the volume of investment into technologies with marginal costs in the range $c$ to $c + \Delta c$. The associated investment cost is $k(c) q'(c) \Delta c$. The total investment cost accounts for all such incremental costs up to $\overline{c}$ (the highest marginal cost for which there is investment).

Second, the social planner in second stage minimizes production cost by starting the cheapest production plants, for which the total output equals the shock $\varepsilon$ and subject to the total installed production capacity $\hat{\varepsilon} = q(\overline{c})$. It can be shown that the expected total production cost plus system cost is:[8]

$$T = \int_0^{\hat{\varepsilon}} f(\varepsilon) C(\varepsilon) d\varepsilon + C(\hat{\varepsilon})(1 - F(\hat{\varepsilon})) + M(\hat{\varepsilon}). \tag{2}$$

The expected production cost has two parts: the first represents outcomes without rationing; the second represents those when forced rationing occurs and production is at full capacity $\hat{\varepsilon}$. Finally, $M(\hat{\varepsilon})$ is the system cost at total production capacity $\hat{\varepsilon}$.

Third, the expected benefits to electricity consumers in the second stage can be calculated in a similar way:[9]

$$B = p^* \int_0^{\hat{\varepsilon}} f(\varepsilon) \varepsilon d\varepsilon + p^* \hat{\varepsilon} (1 - F(\hat{\varepsilon})). \tag{3}$$

[8] Note that $\int_{\hat{\varepsilon}}^{\overline{\varepsilon}} f(\varepsilon) C(\hat{\varepsilon}) d\varepsilon = C(\hat{\varepsilon})(1 - F(\hat{\varepsilon}))$.

[9] Note that $\int_{\hat{\varepsilon}}^{\overline{\varepsilon}} f(\varepsilon) \hat{\varepsilon} d\varepsilon = \hat{\varepsilon}(1 - F(\hat{\varepsilon}))$.

The social planner chooses the distribution function $q(c)$ and technology cutoff $\overline{c}$ to maximize social welfare $W \equiv B - T - K$. We obtain the following result:

**Proposition 1** *It is socially-optimal to make generation investments such that the inverse marginal cost curve becomes:*

$$\begin{aligned} q(c) &= F^{-1}\left(1 + k'(c)\right) \text{ for } c \in [0, \overline{c}] \\ q(c) &= q(\overline{c}) \text{ for } c \in [\overline{c}, p^*], \end{aligned} \tag{4}$$

*where the technology cutoff* $\overline{c}$ *is implicitly determined from:*

$$-(p^* - \overline{c})\, k'(\overline{c}) - k(\overline{c}) - M'(q(\overline{c})) = 0, \tag{5}$$

*which has a unique solution in the range* $[0, p^*]$.

Proposition 1 characterizes socially-optimal investment based on the trade-off between investment and production costs. The condition for optimal investments can be understood by rearranging (4) to give $1 - F(q(c)) = -k'(c)$. Consider a planner choosing between two technologies with a (small) marginal-cost difference $\Delta c$. Investing in the technology with lower marginal cost saves $[1 - F(q(c))]\,\Delta c$ in expected production costs, as $1 - F(q(c))$ is the probability that demand will be larger than $q(c)$, i.e. the probability that the plant will be used. On the other hand, this incurs an extra $[-k'(c)]\,\Delta c$ in investment costs. At the optimum, the social planner is indifferent at the margin between such similar investments. Our optimality condition is simpler but results with a similar intuition have been found in a continuous investment framework (Zöttl 2010) and for discrete peak-load pricing (Crew & Kleindorfer 1976).

A key insight from our first-order condition is that neither the VOLL of consumers $p^*$ nor the system cost $M(\cdot)$ have any influence on socially-optimal investments *below* the technology cutoff $\overline{c}$ —though they do affect the cutoff itself. To explore the cutoff condition further, note that (4) holds also at $\overline{c}$, so we can rewrite the condition as:

$$(p^* - \overline{c})\left(1 - F(q(\overline{c}))\right) - M'(q(\overline{c})) - k(\overline{c}) = 0. \tag{6}$$

The first term gives the consumer benefit from additional investment, and the second term represents the benefit of a lower system cost. The planner continues to invest until these marginal benefits at the optimum $\overline{c}$ are equal to the investment cost $k(\overline{c})$. A higher consumer reservation price (higher $p^*$) and a steeper decrease in system costs (higher $-M'(\cdot)$) both induce more investment into peaking plant.

Optimality can be expressed in another intuitive form. Let $\eta = [-ck'(c)/k(c)]_{c=\overline{c}} > 0$ denote the elasticity of investment costs with respect to production costs, evaluated at the optimum technology cutoff $\overline{c}$. This is a measure of the technology trade-off: $\eta$ is larger if a lower production cost comes with a greater increase in investment cost, so technology is less flexible in that sense. This allows us to rewrite (4):[10]

$$\overline{c} = p^* \left[ \frac{\eta}{\eta + 1 - \dfrac{-M'\left(F^{-1}\left(1 + k'\left(\overline{c}\right)\right)\right)}{k\left(\overline{c}\right)}} \right] < p^*. \tag{7}$$

We see that the cutoff $\overline{c}$ is higher for higher consumer VOLL (higher $p^*$) and less flexible available generation technology (higher $\eta$); the former makes investment more valuable, and the latter makes it more necessary.[11] The cutoff is also higher with a steeper decrease in system costs (higher $-M'(\cdot)$), as this also makes investment more valuable.

In the special case without a system-cost externality ($M(\cdot) \equiv 0$), the cutoff $\overline{c} = p^*\eta/(\eta+1)$ is *independent* of the distribution of consumer demand $F(\cdot)$. As any changes in the shape of $F(\cdot)$ leave $\overline{c}$ unchanged, it follows directly from (4) that the socially-optimal LOLP is also unchanged.

# 4 Market-based investment and capacity mechanisms

We now turn to the investment and technology mix delivered by a competitive market. The regulator has two instruments: setting a price cap in the wholesale electricity market and designing a uniform market-wide capacity payment. Our main interest lies in deriving the optimal policy design that delivers the social optimum.

## 4.1 Model setup and additional assumptions

In Stage 2, produced electricity is paid a spot price $p(\varepsilon)$. In a competitive market, the price can be implicitly determined from $\varepsilon = q(p)$, that is, demand equals supply. Moreover, there is a price cap $\overline{p}$ which the highest spot price allowed by the wholesale market design. In the case of forced demand rationing, the spot price equals the price cap. Our analysis allows for price caps above the VOLL, and we define $\widehat{p} = \min(p^*, \overline{p})$.[12] In Stage 1, producers are paid a "US-style" uniform capacity payment $z \in [0, k(\widehat{p}))$ for each unit of invested capacity.

[10] Recall that our assumptions ensure that $-M'(\cdot)/k(\overline{c}) < 1$ so that $\overline{c} < p^*$.

[11] As the available technology becomes very flexible with $\eta \to 0$, the trade-off between production cost and investment cost disappears, so the planner can achieve first-best while relying almost only on abundant very low marginal cost generation and so the technology cutoff $\overline{c} \to 0$.

[12] As will become clear, when investments are a public good (via the system externality $M(\hat{\varepsilon})$), it can be socially-optimal to occasionally have prices above the VOLL. It is therefore implicit in our setup that consumers do not quit the market so that the socially-optimal solution can be implemented. This could be justified, for example, by such consumers being compensated via the wider tax system.

## 4.2 Competitive equilibrium and optimal policy design

In a competitive market, $q(c)$ is the market supply curve, i.e., the market capacity with marginal cost below $c$. In equilibrium, price equals marginal cost, $p(q(c)) = c$, so a plant with marginal cost $c$ will produce for $\varepsilon \geq q(c)$ which corresponds to $p(\varepsilon) \geq c$.

The expected profit from an investment into a unit of a technology with marginal cost $c \in (0, \overline{c})$ is therefore given by:

$$\pi(c) = z - k(c) + \int_{q(c)}^{\hat{\varepsilon}} (p(\varepsilon) - c) f(\varepsilon) d\varepsilon + (\overline{p} - c)[1 - F(\hat{\varepsilon})]. \tag{8}$$

The first two terms are the capacity payment and investment cost; the third is the profit flow from the spot market. The last term is often referred to as the expected scarcity rent (Stoft, 2002). Competitive entry means that the zero-profit condition $\pi(c) \equiv 0$ holds, for every technology $c \in (0, \overline{c})$, in equilibrium.

Our next result derives the policies that achieve the social optimum via the market:

**Proposition 2** *Investments are socially optimal when i) the price cap is at or above the socially optimal technology cutoff, i.e.* $\overline{p} \geq \overline{c}$*, and ii)*

$$z = -M'(q(\overline{c})) - k'(\overline{c})(p^* - \overline{p}). \tag{9}$$

*This for example holds when the price cap equals the VOLL,* $\overline{p} = p^*$*, and the capacity payment internalizes the marginal system-cost externality,* $z = -M'(q(\overline{c}))$*.*

Proposition 2 formalizes how market-based investment, augmented by an optimally-designed price cap and capacity payment, can replicate the socially-optimal technology mix of Proposition 1.

We derive (9) as the key condition which characterizes the family $(\overline{p}, z)$ of price caps and capacity payments that achieves the social optimum. Figure 2 illustrates. A leading special case is setting the price cap at the VOLL, $\overline{p} = p^*$, and the capacity payment to reflect the marginal system cost, $z = -M'(q(\overline{c}))$. From a policy perspective, this establishes a rationale for the use of a capacity mechanism even where the wholesale market design has the "correct" price cap set at the VOLL. The reason is that an additional instrument is needed to correct for the system-cost externality of generation capacity.[13]

The same condition also makes precise how a capacity mechanism can correct for the "missing money problem". Specifically, suppose that for reasons of political economy, the price cap is set too low with $\overline{p} < p^*$; (9) tells us that a capacity payment of $z = -M'(q(\overline{c})) + (p^* - \overline{p})[-k'(\overline{c})]$ restores optimality. This includes an "uplift" equal to

[13] Fabra (2018) obtains a similar finding with a single generation technology and a price cap always set at the VOLL, $\overline{p} = p^*$. See also Llobet & Padilla (2018) for a related result with two generation technologies. Our analysis goes further by characterizing the family of all optimal policy combinations.

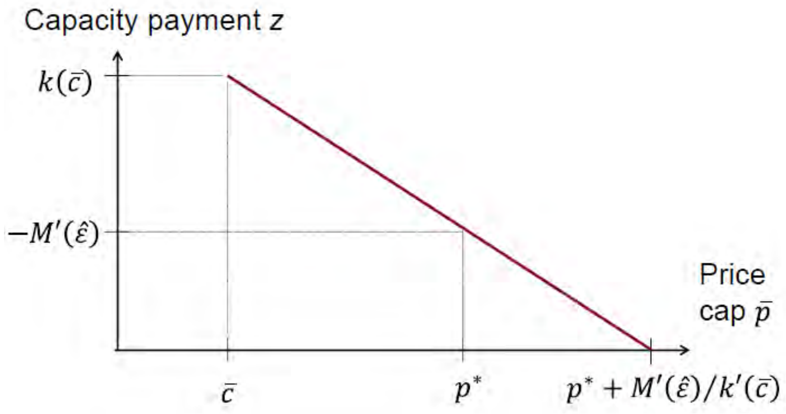


Figure 2: Illustration of the family $(\overline{p}, z)$ of price caps and capacity payments that achieves the social optimum in a competitive market.

$(p^* - \overline{p})\left[-k'(\overline{c})\right]$ relative to full scarcity pricing. From the viewpoint of producers, the split of revenues between the energy and capacity markets can vary widely—even where overall regulation is optimally-designed.

### 4.3 Policy equivalence: Capacity payment vs capacity auction

In our model, it is equivalent for the regulator to set a capacity payment $z$ resulting in market-based invested capacity $\widehat{\varepsilon}$ or to instead set a capacity level $\widehat{\varepsilon}$ which is delivered by a market-based capacity price $z$ in capacity auction.[14] The reason is that producers have symmetric information, so even if there is a demand shock in the spot market, there are no surprises in the capacity market; in equilibrium, producers can predict the outcome of the capacity auction. With symmetric information and small price-taking producers, it is equivalent for the regulator to set "price" (capacity payment) or "quantity" (capacity volume)—akin to Weitzman (1974).

Before deriving comparative statics on the impacts of the two different policy implementations, we want to establish that our results do extend to cases where $(\overline{p}, z)$ do not necessarily belong to the family that achieves the social optimum. First, we establish a generalised version of the technology cutoff. For any price $p > p_0^*$, we can implicitly define a cutoff $\overline{c}(p)$ from

$$-\left(p - \overline{c}(p)\right) k'\left(\overline{c}(p)\right) - k\left(\overline{c}(p)\right) - M'\left(q\left(\overline{c}(p)\right)\right) = 0,$$

which is a generalised version of (5). The results in Proposition 2 can now be generalised as follows.

[14]The cutoff condition in Proposition 1 depend on the VOLL $p^*$ which, in practice, can be difficult to estimate. Cramton and Stoft (2005) recommend that the regulator sets a sufficiently high capacity level $\widehat{\varepsilon}$ to makes LOLP acceptably low—and then procures this using a capacity auction.

**Proposition 3** *For the price cap* $\overline{p} > p_0^*$ *and capacity payment* $z = -M'\left(q\left(\bar{c}\left(\overline{p}\right)\right)\right)$*, the technology cutoff for investments is* $\bar{c}\left(\overline{p}\right)$ *and the supply function is*

$$q\left(c\right) = F^{-1}\left(1 + k'\left(c\right)\right) \;\; \textit{for } c \in \left[0, \bar{c}\left(\overline{p}\right)\right]. \tag{10}$$

*The same investments would follow for an alternative price cap* $\tilde{p} \geq \bar{c}\left(\overline{p}\right)$*, if*

$$\begin{aligned} z &= -M'\left(q\left(\bar{c}\left(\overline{p}\right)\right)\right) - k'\left(\bar{c}\left(\overline{p}\right)\right)\left(\overline{p} - \tilde{p}\right) \\ &= k\left(\bar{c}\left(\overline{p}\right)\right) + k'\left(\bar{c}\left(\overline{p}\right)\right)\left(\tilde{p} - \bar{c}\left(\overline{p}\right)\right). \end{aligned} \tag{11}$$

We will be using these conditions in the comparative statics analysis. Note that somewhat similar to our analysis of the social optimum, the equilibrium condition in (10) implies that market-based investments below the technology cutoff $\bar{c}$ do not depend on the price cap $\overline{p}$ or on the capacity payment $z$. Hence, these policy instruments influence only investments into peaking plant.

### 4.3.1 Comparative statics: Capacity payment and price cap

**Proposition 4** *Consider a capacity mechanism where the capacity payment* $z$ *is set by the market design:*
*(i) The technology cutoff* $\bar{c}$ *increases for a higher capacity payment* $z$*:*

$$\frac{\partial \bar{c}}{\partial z} = \frac{1}{\left(\overline{p} - \bar{c}\right) k''\left(\bar{c}\right)} > 0 \tag{12}$$

*and for a higher price cap* $\overline{p}$*:*

$$\frac{\partial \bar{c}}{\partial \overline{p}} = \frac{-k'\left(\bar{c}\right)}{\left(\overline{p} - \bar{c}\right) k''\left(\bar{c}\right)} > 0. \tag{13}$$

*(ii) The loss of load probability,* $1 - F\left(q\left(\bar{c}\right)\right)$*, decreases for a higher capacity payment* $z$*:*

$$\frac{\partial\left(1 - F\left(q\left(\bar{c}\right)\right)\right)}{\partial z} = -f\left(q\left(\bar{c}\right)\right) q'\left(\bar{c}\right) \frac{\partial \bar{c}}{\partial z} < 0$$

*and decreases for a higher price cap* $\overline{p}$*:*

$$\frac{\partial\left(1 - F\left(q\left(\bar{c}\right)\right)\right)}{\partial \overline{p}} = -f\left(q\left(\bar{c}\right)\right) q'\left(\bar{c}\right) \frac{\partial \bar{c}}{\partial \overline{p}} < 0.$$

Proposition 4 shows that a higher capacity payment and price cap both raise investment and decrease the LOLP. In this sense, these two policy instruments are substitutes.

### 4.3.2 Comparative statics: Capacity volume and price cap

**Proposition 5** *Consider a capacity mechanism where the capacity volume $\hat{\varepsilon}$ is set by the market design and the capacity payment $z$ is endogenously determined. A higher procured capacity $\hat{\varepsilon}$ raises the technology cutoff $\overline{c}$:*

$$\frac{\partial \overline{c}}{\partial \hat{\varepsilon}} = \frac{f(\hat{\varepsilon})}{k''(\overline{c})} > 0, \tag{14}$$

*and raises the capacity payment $z$:*

$$\frac{\partial z}{\partial \hat{\varepsilon}} = (\overline{p} - \overline{c}) f(\hat{\varepsilon}) > 0. \tag{15}$$

Proposition 5 shows how a capacity auction that procures more capacity investment brings a higher technology cutoff and requires a higher capacity payment. The first part of the result, $\partial \overline{c}/\partial \hat{\varepsilon} = f(\hat{\varepsilon})/k''(\overline{c}) > 0$, follows directly from the condition (10) which ensures that producers are indifferent between investment alternatives.

The second part of the result can be understood as follows. If investments increase by a (small) $\Delta\widehat{\varepsilon}$, producers will now be paid $\overline{c}$ instead of the higher price cap $\overline{p}$ for shocks in the range $[\hat{\varepsilon}, \hat{\varepsilon} + \Delta\widehat{\varepsilon}]$. Shocks are in this range with probability $f(\hat{\varepsilon})\Delta\widehat{\varepsilon}$. Therefore, to ensure that the expected profit from marginal investments remains zero, a competitive market adjusts the capacity payment upwards by $(\overline{p} - \overline{c}) f(\hat{\varepsilon})\Delta\widehat{\varepsilon}$.

**Proposition 6** *Consider a capacity mechanism where the capacity volume $\hat{\varepsilon}$ is set by the market design and the capacity payment $z$ is endogenously determined. For a given procured capacity $\hat{\varepsilon}$, a higher price cap $\overline{p}$ has no impact on the technology cutoff $\overline{c}$:*

$$\left.\frac{d\overline{c}}{d\overline{p}}\right|_{\hat{\varepsilon}\ \mathit{fixed}} = 0, \tag{16}$$

*and reduces the capacity payment $z$:*

$$\left.\frac{dz}{d\overline{p}}\right|_{\hat{\varepsilon}\ \mathit{fixed}} = k'(\overline{c}) < 0. \tag{17}$$

The first part of Proposition 6 again follows directly from (10); the technology cutoff is unchanged if invested capacity is unchanged. The second part can be understood by using $[1 - F(\hat{\varepsilon})] = -k'(\overline{c})$ from (4) to rewrite (17) as $[dz/d\overline{p}]_{\hat{\varepsilon}\ \text{fixed}} = -[1 - F(\hat{\varepsilon})]$. Producers are paid the price cap $\overline{p}$ when there is demand rationing, which occurs with probability $1 - F(\hat{\varepsilon})$. Hence, if the price cap is increased by a (small) $\Delta\overline{p}$, then the capacity payment must decline by $\Delta\overline{p}[1 - F(\hat{\varepsilon})]$ to keep expected profit from marginal investments at zero.

### 4.3.3 Payoff impacts of a capacity mechanism and price cap

How does a capacity mechanism affect the overall payoff of producers? On one hand, by Proposition 4, a higher capacity payment makes outcomes with forced rationing less likely, which reduces producers' expected scarcity rent. On the other hand, a higher capacity payment creates an additional revenue stream. Similarly, a higher price cap means more revenue in the event of forced rationing but also makes this event less likely. Our next result clarifies the net effect of these two forces:

**Proposition 7** *Consider a market with the initial technology cutoff* $\overline{c}_0$*. If the capacity payment* $z$ *and/or the procured capacity* $\widehat{\varepsilon}$ *and/or the price cap* $\overline{p}$ *is increased, then:*
*(i) Total expected revenue from the spot market and the capacity mechanism is unchanged for each plant with a marginal cost below the cutoff* $\overline{c}_0$*.*
*(ii) All of the additional expected revenue covers production and investment costs of new generation investments into plant with a marginal cost above* $\overline{c}_0$*.*

Proposition 7 shows that, in equilibrium, a higher capacity payment or price cap have no impact (neither in terms of payoffs or investments) on plant below the technology cutoff. This is consistent with our earlier finding from Proposition 2 that these do not influence investment below the technology cutoff. Given their equivalence, the same conclusion also applies to higher procured capacity. The resulting additional revenues to producers solely finance the new generation investment occurring at/above the old technology cutoff.

# 5 The impact of renewables penetration

A central feature of the future electricity market is that it will be dominated by renewable generation from solar and wind. In the policy debate, the growth of intermittent renewables and its adverse impact on the demand for conventional generation is frequently asserted as a justification for a capacity mechanism. In this section, we use our model to formally characterize the impact of renewables on investment in conventional generation, the socially-optimal LOLP and on the optimal design of capacity mechanism.

## 5.1 Model setup and additional assumptions

We generalize the model as follows. First, let $w$ denote the (exogenous) level of installed renewables capacity, interpreted as consumers' own production such as rooftop solar power or other intermittent renewable generation such as wind power supplied by an exogenous fringe (with zero marginal cost). Write $F(\varepsilon, w)$ as the probability distribution of *net* demand for conventional plant and assume that the probability that net demand is below some level $\varepsilon$ increases with more renewable generation, $\frac{\partial F(\varepsilon;w)}{\partial w} > 0$. We allow the strength of the crowding-out effect to vary along $F(\cdot)$ (e.g., with the time of year). Second, write $M(\hat{\varepsilon}, w)$ as the system-cost externality, where $\hat{\varepsilon}$ is installed conventional capacity. We assume that the system cost rises with renewables, $\frac{\partial M(\hat{\varepsilon},w)}{\partial w} \geq 0$, and that this effect is mitigated by conventional generation, $\frac{\partial}{\partial \hat{\varepsilon}}\left[\frac{\partial M(\hat{\varepsilon},w)}{\partial w}\right] = \frac{\partial^2 M(\hat{\varepsilon},w)}{\partial \hat{\varepsilon} \partial w} \leq 0$. The former describes how intermittent renewables make it more difficult to control the power system; the latter captures the idea that "firm capacity" acts as complement to this intermittency.[15] Third, we use $q(c, w)$ to denote socially-optimal conventional supply with marginal cost below $c$, so $q(c, w) = \varepsilon$ is the condition for market clearing given renewables $w$.

## 5.2 Renewables and optimal investment

The optimality conditions from Proposition 1 for investment and the technology cut-off from (4) and (5) are valid for any net-demand distribution $F$. It does not matter whether that probability distribution has been influenced by wind power or not. Hence, for installed wind capacity $w$, the condition for optimal investment from (4) becomes:

$$1 - F(q(c, w), w) = -k'(c). \tag{18}$$

Similarly, the condition for the optimal technology cutoff $\bar{c}$ from (5) becomes:

$$-(p^* - \bar{c})\, k'(\bar{c}) - k(\bar{c}) - \frac{\partial M(q(\bar{c}, w), w)}{\partial \hat{\varepsilon}} = 0, \tag{19}$$

where $q(\bar{c}, w) = \hat{\varepsilon}$ is the total conventional production capacity and the LOLP is $1 - F(\hat{\varepsilon}, w)$.

This leads immediately to the following initial finding:

**Lemma 1** *For a given technology mix (i.e., fixed $\bar{c}$), higher renewables capacity reduces conventional supply, $\frac{\partial q(c,w)}{\partial w} < 0$.*

For a given technology mix, higher renewable capacity always reduces the conventional supply. This is an instance of the widely-discussed merit-order effect of renewables penetration. It is a consequence of more renewables making low realizations of net demand for conventional plant more likely ($\frac{\partial F(\varepsilon;w)}{\partial w} > 0$).

Our first result builds on this to characterize the equilibrium impact of more renewables capacity on the optimal technology cutoff and LOLP:

**Proposition 8** *Higher renewables capacity raises the optimal technology cutoff, $\frac{d\bar{c}}{dw} \geq 0$, and reduces the socially-optimal LOLP $\frac{dLOLP}{dw} \leq 0$. In the special case with no system-cost externality, i.e., $M(\hat{\varepsilon}, w) \equiv 0$, both impacts are zero, $\frac{d\bar{c}}{dw} = 0$ and $\frac{dLOLP}{dw} = 0$.*

[15] For a given level of $w$, we retain our previous assumptions on $F(\cdot)$ and $M(\cdot)$.

The main insight from Proposition 8 is that higher renewables capacity raises the socially-optimal technology cutoff. In other words, it becomes optimal to bring to market some conventional plant technologies with high marginal cost that were previously not needed. This formalizes the commonly-expressed view that renewables raise the social value of peaking plant. An immediate implication is that, at the social optimum, the LOLP declines with more renewables.

These renewables impacts hinge crucially on the presence of the system-cost externality $M(\hat{\varepsilon}, w)$. They also rely on the complementarity property: having more conventional plant mitigates any incremental system externality created by renewables. In its absence, more renewables capacity has zero impact on the social optimum. This follows directly from Proposition 1: if $M(\cdot) \equiv 0$, optimal investment and the technology mix are independent of the distribution of net demand.[16]

Note the tension underlying these findings. On one hand, the "merit-order effect" from Lemma 1 is that, for a given technology mix $\bar{c}$, renewables reduce conventional supply. On the other hand, the "system-complementarity effect" from Proposition 8 says that renewables raise the technology cutoff $\bar{c}$, so additional peaking plant are needed. Our next result presents a condition on the overall impact on conventional capacity:

**Proposition 9** *Higher renewables capacity reduces the socially-optimal conventional capacity, $\frac{dq(\bar{c},w)}{dw} \leq 0$, if the merit-order effect dominates the system-complementarity effect:*

$$\left[(p^* - \bar{c}) - \frac{\partial^2 M(\hat{\varepsilon}, w)}{\partial \hat{\varepsilon}^2} \Big/ \frac{\partial F(\hat{\varepsilon}, w)}{\partial \varepsilon}\right] \frac{\partial F(\hat{\varepsilon}, w)}{\partial w} \geq -\frac{\partial^2 M(\hat{\varepsilon}, w)}{\partial \hat{\varepsilon} \partial w}.$$

In general, the overall impact, taking into account the knock-on effect of renewables on the optimal technology mix, is theoretically ambiguous. Proposition 9 makes precise when the overall impact is such that the socially-optimal conventional generation capacity declines. The underlying condition is that the residual-demand effect of more renewables dominates the system-complementarity effect, that is, the cross-partial $\frac{\partial^2 M(\hat{\varepsilon},w)}{\partial \hat{\varepsilon} \partial w}$ cannot be too negative.[17]

## 5.3 Renewables and capacity-mechanism design

We now turn to characterizing the impact of more renewables on the design of a capacity mechanism. Like before, we begin by noting that the result from Proposition 2 on the price cap and capacity payment that achieve social optimality remains valid for any given

[16] Biggar & Hesamzadeh (2014, pp. 192-194) obtain an instance of this finding from a graphical screening curve analysis with two conventional technologies.

[17] The condition is, of course, also always met in the absence of any system-cost externality, $M(\cdot) \equiv 0$.

$w$. In particular, the family of socially-optimal policy instruments from (9) now becomes:

$$z\left(w, \overline{p}\right) + \frac{\partial M\left(q\left(\overline{c}, w\right), w\right)}{\partial \hat{\varepsilon}} + \left(p^{*} - \overline{p}\right) k'\left(\overline{c}\right) = 0, \tag{20}$$

which makes explicit the potential dependency of the capacity payment $z\left(w, \overline{p}\right)$ on renewables and the price cap.

Our next result formalizes how higher renewables penetration affects the design of the capacity payment:

**Proposition 10** *For any price cap $\overline{p}$, higher renewables capacity increases the socially-optimal capacity payment, $\frac{\partial z(w,\overline{p})}{\partial w} \geq 0$. In the special case with no system-cost externality, i.e., $M(\hat{\varepsilon}, w) \equiv 0$, the impact is zero, $\frac{\partial z(w,\overline{p})}{\partial w} = 0$.*

Proposition 10 shows how increased renewables penetration exacerbates the need for a capacity mechanism. For any given level of the price cap, the socially-optimal optimal capacity payment increases. Hence the optimal family of policy instruments $(\overline{p}, z)$ shown in Figure 2 is pushed outwards.

This is a direct implication of our finding from Proposition 8 that renewables increase the social value of peaking plant. Given that the price cap is held constant, this additional investment is optimally procured by way of a higher capacity payment. Once again, the result hinges crucially on the presence of the system-cost externality.

In sum, our model shows how the combination of increased renewables penetration and a system-cost complementarity between renewables and conventional plant raises the social value of peaking plant and can justify higher capacity payments to conventional generators.

# 6 An optimally-designed strategic reserve

Some countries use a strategic reserve instead of a market-wide capacity market. A reserve is discriminatory in that a capacity payment is made only to generation units within the reserve. An argument in favour is that this limits the market operations of the system operator (SO) to procuring a (small) reserve. This is an advantage in Europe where a SO often owns the transmission network and accordingly has congestion rents—and the regulator wishes to contain this dominant position. It is less of an issue in restructured US electricity markets with independent SOs (ISOs) that do not own any grid assets.[18]

Yet it is also clear that discriminating between plants inside and outside the reserve can easily lead to market distortions. For example, if plants both inside and outside the reserve are paid the same electricity price when the reserve in used, which is the case in

[18] This also explains why US day-ahead markets are centralized and organized by an ISO while markets are more decentralized in Europe (Ahlquist et al. 2019).

Sweden, then the revenue of plants in the reserve are disproportionately large, as they also get a capacity payment. This distorts investments.

We next present an optimally-balanced market design with a strategic reserve that avoids any such competitive inefficiencies. In a nutshell, this can be achieved by paying an extra-high spot price to non-reserve plants whenever the reserve is used.

## 6.1 Model setup and additional assumptions

The model is a variation on the previous setup. In Stage 1, plants in the strategic reserve is paid a uniform capacity payment $z \in [0, k(\widehat{p}))$ for each unit of invested capacity. In Stage 2, electricity produced outside the reserve is paid a spot price $p(\varepsilon)$. We let $q(p)$ and $\widehat{\varepsilon}$ denote the supply and total capacity of non-reserve plants, respectively. Hence, for $\varepsilon \in (0, \widehat{\varepsilon})$, the spot price can be implicitly determined from $\varepsilon = q(p)$. The strategic reserve is "triggered" when the non-reserve capacity has been exhausted, i.e., $\varepsilon > \widehat{\varepsilon}$. In this case, the spot price for non-reserve plant is at the price cap $\overline{p}$—irrespective of whether there is demand rationing. This means revenues of non-reserve plants are independent of the size of the strategic reserve. We let $\widehat{\varepsilon}_r$ denote total production capacity including the reserve.

In a competitive market, reserve plant bid and offer at marginal cost. We let $q_r(c)$ be the total market supply curve, including supply from the reserve. If the reserve is used, $\varepsilon > \widehat{\varepsilon}$, but there is no forced rationing, $\varepsilon < \widehat{\varepsilon}_r$, then the clearing price of the reserve $p_r$ is determined from $\varepsilon = q_r(p_r)$. We have $p_r < \overline{p}$, so for the demand range $\varepsilon \in (\widehat{\varepsilon}, \widehat{\varepsilon}_r)$, plants in the reserve are paid a lower spot price than plants outside the reserve. If the reserve is used and there is forced rationing, $\varepsilon > \widehat{\varepsilon}_r$, also plants in the reserve are paid the price cap $\overline{p}$. Let $\bar{c}$ be the highest marginal cost for which there is investment in the conventional market, and let $\bar{c}_r$ be the highest marginal cost for which there is investment in the reserve.

## 6.2 Competitive equilibrium and optimal policy design

Competitive entry ensures that the zero-profit condition $\pi(c) \equiv 0$ holds in equilibrium, both for plants in the reserve and outside the reserve. Under the above assumptions, we can show that the strategic reserve is equivalent to a US-style market with a uniform capacity payment $z$ and price cap $\overline{p}$:

**Proposition 11** *For a market design with a strategic reserve where* $\overline{p} > p_0^* + M'(q(\bar{c}(p_0^*))) / k'(\bar{c}(p_0^*))$*:*[19]

[19] Recall that $p_0^*$ is the lowest VOLL level for which we can ensure positive socially optimal investments. Hence, $\bar{c}(p_0^*)$ and $q(\bar{c}(p_0^*))$ are the socially-optimal technology cutoff and market capacity, respectively, for that lowest VOLL level.

(i) There is a highest marginal cost $\overline{c} \in (0, \widehat{p})$ for which there is investment in the non-reserve market, where this cutoff satisfies:

$$-k(\overline{c}) - (\overline{p} - \overline{c})\, k'(\overline{c}) = 0. \tag{21}$$

(ii) There is a highest marginal cost $\overline{c}_r > \overline{c}$ for which there is investment in the reserve, where this cutoff is determined from the same condition as for a uniform capacity payment $z$ and a price cap $\overline{p}$:

$$z - k(\overline{c}_r) - (\overline{p} - \overline{c}_r)\, k'(\overline{c}_r) = 0. \tag{22}$$

(iii) Investments in the non-reserve and the reserve give rise to a total supply curve $q_r(c)$, which below the cutoff $\overline{c}_r$ is determined by:

$$q_r(c) = F^{-1}(1 + k'(c)). \tag{23}$$

(iv) Investments are socially optimal whenever the price cap $\overline{p}$ and the capacity payment $z$ to the reserve satisfy:

$$z + M'(q(\overline{c}_r)) + (p^* - \overline{p})\, k'(\overline{c}_r) = 0, \tag{24}$$

which for example holds if the price cap equals the VOLL, $\overline{p} = p^*$, and the capacity payment to the reserve internalizes the marginal system-cost externality, $z = -M'(q(\overline{c}_r))$.
(v) When the reserve is used, a non-reserve plant is paid a higher spot price than reserve plant. The difference $\overline{p} - p_r$ is, in expectation, equal to the capacity payment $z$ of the reserve.

Part (v) is the central result of Proposition 11: it is possible to design a targeted strategic reserve that is as distortion-free as a market-wide capacity payment. As a whole, the market design with a strategic reserve does not discriminate between plants inside and outside the reserve. This non-discrimination property is crucial to avoiding over- or underinvestment in either conventional plants or in the strategic reserve itself.

The optimal design requires that, whenever the strategic reserve is used, the spot price should be at the price cap, while plants in the reserve should be paid the clearing price of the reserve. Given this, the strategic reserve is as efficient as the market design from Proposition 2, with an identical price cap and a discriminatory capacity payment to the strategic reserve at the same level as the previous market-wide capacity payment.

Parts (i)–(iv) are analogs to now familiar conditions from Proposition 2.

As far as we know, this optimal design of a strategic reserve is novel. It has both similarities and differences relative to how strategic reserves are operated in practice. The underlying principle of trying to isolate the operation of the reserve from the wholesale market appears to be well-understood. A central feature of our design is that reserve plant—in addition to receiving a capacity payment—make competitive bids so there is also a clearing price for the reserve itself. In this sense, our design captures symmetrically the benefits of competition for both non-reserve and reserve plant.

# 7 Conclusions and policy implications

We have introduced a new benchmark model of long-run investment and the optimal design of a capacity mechanism in a competitive electricity market. Relative to existing literature, the main differentiating features of our approach are: (i) a continuum of generation technologies which represents the range of baseload, mid-merit and peaking plant (ii) joint modelling, also as a continuum, of two policy instruments: a wholesale price cap and capacity payment, and (iii) an externality arising from the system costs associated with a blackout.

We showed how socially-optimal generation investment can be achieved through the market using different combinations of a price cap and a capacity payment (or, equivalently, a capacity auction). From a policy perspective, we obtain a rationale for the use of a capacity mechanism even where the wholesale-market design has the "correct" price cap set at the VOLL. The reason is that an additional instrument is needed to correct for a system-cost externality that arises in the event of a black-out. Our characterization of the family of optimal policies makes precise how much "uplift" in a capacity payment is needed to correct for different degrees of missing money.

We presented two extensions. First, we showed how increased penetration of intermittent renewables exacerbates the need for a capacity mechanism. In the presence of a system-cost complementarity between renewables and conventional generation, more renewables raise the social value of peaking plant and thereby justify a higher capacity payment. Second, we outlined a new socially-optimal design of a strategic reserve with a discriminatory capacity payment. It avoids investment distortions between plants inside and outside the reserve by paying an extra-high spot price to non-reserve plant and thereby making the non-reserve market independent of the size of the strategic reserve.

More broadly, our analysis suggests that capacity mechanisms may be a longer-term feature of optimal electricity market design, especially in the presence of high renewables penetration, rather than merely being a fix to a near-term supply crunch. Future research may wish to build on our benchmark results to study the implications of the exercise of market power.

# References

Ahlqvist, V., P. Holmberg, and T. Tangerås (2019). "Central- versus Self-Dispatch in Electricity Markets.", IFN Working Paper 1257, Research Institute of Industrial Economics, Stockholm.

Biggar, D.R. and M. R. Hesamzadeh (2014). *The Economics of Electricity Markets.* IEEE Press.

Cramton, P. and S. Stoft (2005). "A Capacity Market that Makes Sense." *The Electricity Journal*, 18(7): 43–54.

Crew, M. and P. Kleindorfer (1976). "Peak Load Pricing with a Diverse Technology." *Bell Journal of Economics* 7(1): 207–231.

Crew, M. and P. Kleindorfer (1986). *The Economics of Public Utility Regulation.* MIT Press.

EC (2016). Final Report of the Sector Inquiry on Capacity Mechanisms. European Commission, SWD (2016) 385 final, Brussels, 30 November 2016.

Fabra, N. (2018). "A Primer on Capacity Mechanisms." *Energy Economics* 75: 323–335.

Joskow, P.L. (2008). "Capacity Payments in Imperfect Electricity Markets: Need and Design." *Utilities Policy* 16(3): 159–170.

Joskow, P. and J. Tirole (2007). "Reliability and Competitive Electricity Markets." *RAND Journal of Economics* 38(1): 60–84.

Léautier, T.-O. (2016). "The Visible Hand: Ensuring Optimal Investment in Electric Power Generation." *The Energy Journal* 37(2): 89–108.

Léautier, T.-O. (2019). *Imperfect Markets and Imperfect Regulation: An Introduction to the Microeconomics and Political Economy of Power Markets.* MIT Press.

Llobet, G. and J. Padilla (2018). "Conventional Power Plants in Liberalized Electricity Markets with Renewable Entry." *The Energy Journal* 39(3): 69–91.

Newbery, D. and M. Grubb (2015). "Security of Supply, the Role of Interconnectors and Option Values: Insights from the GB Capacity Auction." *Economics of Energy & Environmental Policy* 4(2): 65–82.

Stoft, S. (2002). *Power System Economics: Designing Markets for Electricity.* IEEE Press.

Tangerås, T. P. (2018). "Equilibrium Supply Security in a Multinational Electricity Market with Renewable Production." *Energy Economics* 72: 416-435.

Teirilä, J. and R.A. Ritz (2019). "Strategic Behaviour in a Capacity Market? The New Irish Electricity Market Design." *The Energy Journal*, forthcoming.

Weitzman, M.L. (1974). "Prices vs Quantities." *Review of Economic Studies* 41(4): 477–491.

Zöttl, G. (2010). "A Framework of Peak Load Pricing with Strategic Firms." *Operations Research* 58(6): 1637–1649.

# Appendix

**Proof of Proposition 1.** The proof for the social optimum proceeds in two main steps. In Step 1, we find the supply function $q(c)$ that minimizes the sum of investment and production costs for a given technology cutoff $\overline{c}$. In Step 2, we find the optimal technology cutoff $\overline{c}$ and corresponding optimal investment level $q(\overline{c})$ that reflect consumer preferences. In the proof, we make use of the function $H(\varepsilon)$, defined such that $H(0) = 0$ and $H'(\varepsilon) = F(\varepsilon)$.

*Step 1.* We wish to find the function $q(c)$ that minimizes the sum of investment cost $K$ and expected production cost $T$ for a given technology cutoff $\overline{c}$ and the corresponding $q(\overline{c})$. This becomes straightforward once we have an expression that involves $q(c)$ but none of its derivatives; then we can find an optimal $q(c)$ for each $c$, independently of other marginal-cost levels. Also, we want to get rid of related terms involving $C(\cdot)$ and $C'(\cdot)$.

First, we rewrite the investment cost in this desired form. Using integration by parts, (1) can be rewritten as:

$$\begin{aligned} K &= \int_0^{\overline{c}} k(c)\, q'(c)\, dc = [q(c)\, k(c)]_0^{\overline{c}} - \int_0^{\overline{c}} q(c)\, k'(c)\, dc \\ &= q(\overline{c})\, k(\overline{c}) - q(0)\, k(0) - \int_0^{\overline{c}} q(c)\, k'(c)\, dc. \end{aligned} \tag{25}$$

Second, we use integration by parts to rewrite the expected production cost (2):

$$\begin{aligned} T &= \int_0^{\hat{\varepsilon}} f(\varepsilon)\, C(\varepsilon)\, d\varepsilon + C(q(\overline{c}))\,(1 - F(\hat{\varepsilon})) + M(\hat{\varepsilon}) \\ &= [F(\varepsilon)\, C(\varepsilon)]_0^{\hat{\varepsilon}} - \int_0^{\hat{\varepsilon}} F(\varepsilon)\, C'(\varepsilon)\, d\varepsilon + C(q(\overline{c}))\,(1 - F(\hat{\varepsilon})) + M(\hat{\varepsilon}) \\ &= C(q(\overline{c})) - \int_0^{\hat{\varepsilon}} F(\varepsilon)\, C'(\varepsilon)\, d\varepsilon + M(\hat{\varepsilon}) \\ &= C(q(\overline{c})) - \int_0^{\overline{c}} F(q(c))\, cq'(c)\, dc + M(\hat{\varepsilon}) \\ &= \int_0^{\overline{c}} c\,(1 - F(q(c)))\, q'(c)\, dc + C(q(0)) + M(\hat{\varepsilon}) \\ &= \int_0^{\overline{c}} c\frac{d}{dc}\,[q(c) - H(q(c))]\, dc + C(q(0)) + M(\hat{\varepsilon}) \\ &= [cq(c) - cH(q(c))]_0^{\overline{c}} - \int_0^{\overline{c}} (q(c) - H(q(c)))\, dc + C(q(0)) + M(\hat{\varepsilon}) \\ &= \overline{c}\hat{\varepsilon} - \overline{c}H(\hat{\varepsilon}) + C(q(0)) - \int_0^{\overline{c}} (q(c) - H(q(c)))\, dc + M(\hat{\varepsilon}). \end{aligned} \tag{26}$$

We wish to minimize $T+K$ which using (25) and (26) is equivalent to:

$$T+K = q(\bar{c})k(\bar{c}) - q(0)k(0) + \hat{\varepsilon}\bar{c} + M(\hat{\varepsilon}) - H(\hat{\varepsilon})\bar{c} + C(q(0)) + \int_0^{\bar{c}} \underbrace{H(q(c)) - q(c) - q(c)k'(c)}_{=L} dc. \tag{27}$$

As the remaining terms do not depend on $c$, we want to find the $q(c)$ that minimizes $L$ for each $c \in [0, \bar{c})$, which can now be done independently of $\bar{c}$ and $q(\bar{c})$. The first- and second-order conditions are:

$$\frac{\partial L}{\partial q} = F(q) - 1 - k'(c) = 0 \text{ and } \frac{\partial^2 L}{\partial q^2} = f(q) \geq 0. \tag{28}$$

The second-order condition ensures that the first-order solution is a global minimum. Hence the first-order condition gives the cost-efficient technology mix.

We can represent this technology mix by the supply function $\hat{q}(c)$, which corresponds to an inverse marginal cost curve. The assumed properties of $k'(c)$ ensure that $F(q) = 1 + k'(c) \in (0,1)$. Hence, invertibility of $F(q)$ over this range and the support of this function, implies that for every $c \in [0, \bar{c})$ we have a unique solution $\hat{q}(c) \geq 0$. To confirm that $\hat{q}(c)$ is monotonic and therefore a valid solution, we differentiate the first-order condition:

$$f(\hat{q})\hat{q}'(c) - k''(c) = 0 \Longrightarrow \hat{q}'(c) = \frac{k''(c)}{f(\hat{q})} > 0. \tag{29}$$

*Step 2.* We now wish to find the optimal technology cutoff $\bar{c}$ and optimal investment level $q(\bar{c})$. Using (3) and (27) expected social welfare $W \equiv B - T - K$ can be written as:

$$\begin{aligned} W = {} & p^* \int_0^{q(\bar{c})} f(\varepsilon)\varepsilon d\varepsilon + p^* q(\bar{c})(1 - F(q(\bar{c}))) \\ & - q(\bar{c})\bar{c} + H(q(\bar{c}))\bar{c} - C(q(0)) - \int_0^{\bar{c}} (H(q(c)) - q(c))\, dc \\ & + q(0)k(0) - q(\bar{c})k(\bar{c}) + \int_0^{\bar{c}} q(c)k'(c)\, dc - M(q(\bar{c})). \end{aligned}$$

Differentiating this expression yields:

$$\begin{aligned}
\frac{\partial W}{\partial \bar{c}} &= p^* q'(\bar{c}) f(q(\bar{c})) q(\bar{c}) + p^* q'(\bar{c}) (1 - F(q(\bar{c}))) \\
&\quad - p^* q(\bar{c}) f(q(\bar{c})) q'(\bar{c}) - q'(\bar{c}) \bar{c} - q(\bar{c}) + \\
&\quad + F(q(\bar{c})) \bar{c} q'(\bar{c}) + H(q(\bar{c})) \\
&\quad - (H(q(\bar{c})) - q(\bar{c})) - q'(\bar{c}) k(\bar{c}) - q(\bar{c}) k'(\bar{c}) \\
&\quad + q(\bar{c}) k'(\bar{c}) - M'(q(\bar{c})) q'(\bar{c}) \\
&= p^* q'(\bar{c}) (1 - F(q(\bar{c}))) - q'(\bar{c}) \bar{c} + F(q(\bar{c})) \bar{c} q'(\bar{c}) \\
&\quad - q'(\bar{c}) k(\bar{c}) - M'(q(\bar{c})) q'(\bar{c}) \\
&= (p^* - \bar{c}) q'(\bar{c}) (1 - F(q(\bar{c}))) - q'(\bar{c}) k(\bar{c}) - M'(q(\bar{c})) q'(\bar{c}) \\
&= q'(\bar{c}) \left( \underbrace{(p^* - \bar{c})(1 - F(q(\bar{c}))) - k(\bar{c}) - M'(q(\bar{c}))}_{=Y(\bar{c})} \right). \qquad (30)
\end{aligned}$$

This implies the first-order condition for social welfare is:

$$Y(\bar{c}) = 0. \qquad (31)$$

We next confirm that $Y(\bar{c})$ is decreasing when the technology mix is efficient, that is, (28) is satisfied:

$$\begin{aligned}
Y'(\bar{c}) &= -(1 - F(q(\bar{c}))) - (p^* - \bar{c}) f(q(\bar{c})) q'(\bar{c}) - k'(\bar{c}) - M''(q(\bar{c})) q'(\bar{c}) \\
&= k'(c) - (p^* - \bar{c}) k''(c) - k'(\bar{c}) - M''(q(\bar{c})) q'(\bar{c}) \\
&< 0 \text{ for } \bar{c} \in [0, p^*],
\end{aligned}$$

where the simplification makes use of the first-order condition (28) and its derivative in (29). The first-order condition (28) and the assumed properties of $k(c)$ and $M(\widehat{\varepsilon})$ imply that:

$$\begin{aligned}
Y(0) &= -p^* k'(0) - k(0) - M'(q(0)) > 0 \\
Y(p^*) &= -k(p^*) - M'(q(p^*)) < 0.
\end{aligned}$$

Together with the property $Y'(\bar{c}) < 0$, this ensures a unique solution to $Y(\bar{c}) = 0$ in the range $(0, p^*)$. Since also $q'(\bar{c}) > 0$, it follows from (30) that:

$$\frac{\partial W}{\partial \bar{c}} \geq 0 \text{ for } c \in [0, \bar{c}] \text{ and } \frac{\partial W}{\partial \bar{c}} \leq 0 \text{ for } c \in [\bar{c}, p^*],$$

so we can conclude that the first-order condition $Y(\bar{c}) = 0$ gives a social (global) optimum.

The following lemma is useful when proving Proposition 2, where we e.g. show that the conditions (32) and (33) are satisfied for the markets that we consider.

**Lemma 2** *Consider a perfectly competitive market, where the capacity payment is not too low or too high, i.e.*

$$z - k(0) - \overline{p}k'(0) > 0 \tag{32}$$
$$z - k(\widehat{p}) - (\overline{p} - \widehat{p})\,k'(\widehat{p}) < 0. \tag{33}$$

*In this case, the following can be proven:*

1. *There is a highest marginal cost* $\overline{c} \in (0, \widehat{p})$ *for which there is investment. This cutoff can be uniquely determined from:*

$$z - k(\overline{c}) - (\overline{p} - \overline{c})\,k'(\overline{c}) = 0. \tag{34}$$

2. *Investments give rise to a supply curve* $q(c)$*, which can be determined from:*

$$q(c) = F^{-1}(1 + k'(c)) \tag{35}$$

*for* $c \in [0, \overline{c}]$.

**Proof of Lemma 2.** With competitive entry, the zero-profit condition $\pi(c) \equiv 0$ is an identity. Hence, since entrants are price takers (so that $p(q(c)) = c$) and using Leibniz' rule, we can differentiate both sides of (8) to get:

$$\begin{aligned}
\pi'(c) &= -k'(c) - q'(c)(p(q(c)) - c) f(q(c)) - \int_{q(c)}^{\hat{\varepsilon}} f(\varepsilon)\, d\varepsilon - (1 - F(\hat{\varepsilon})) \\
&= -k'(c) - (1 - F(q(c))) \\
&= 0,
\end{aligned}$$

which is the first-order condition in (10). For the marginal technology $\overline{c}$, we have $\hat{\varepsilon} = q(\overline{c})$ so the zero-profit condition based on (8) simplifies to:

$$\begin{aligned}
\pi(\overline{c}) &= z - k(\overline{c}) + (\overline{p} - \overline{c})(1 - F(\hat{\varepsilon})) && (36) \\
&= z - k(\overline{c}) - (\overline{p} - \overline{c})\,k'(\overline{c}) && (37) \\
&= 0, && (38)
\end{aligned}$$

where the second step uses (10). For given $z$ and $\overline{p}$, we can use this condition to solve for $\overline{c}$. To show that such a solution exists and is unique, let $\overline{\pi}(\overline{c}) = z - k(\overline{c}) - (\overline{p} - \overline{c})\,k'(\overline{c})$ and differentiate to get:

$$\overline{\pi}'(\overline{c}) = -k'(\overline{c}) + k'(\overline{c}) - (\overline{p} - \overline{c})\,k''(\overline{c}) < 0 \text{ for } \overline{c} \in [0, \widehat{p}).$$

It follows from the conditions in (32) and (33) that:

$$\begin{aligned}\overline{\pi}(0) &= z - k(0) - \overline{p}k'(0) > 0,\\ \overline{\pi}(\widehat{p}) &= z - k(\widehat{p}) - (\overline{p} - \widehat{p})\, k'(\widehat{p}) < 0.\end{aligned}$$

These inequalities and $\overline{\pi}'(\overline{c}) < 0$ ensure a unique solution to $\overline{\pi}(\overline{c}) = 0$ in the range $(0, \widehat{p})$.

**Proof of Proposition 2.** We note that the first-order condition for investments in (35) is identical to that of Proposition (1). Hence, the competitive market yields socially-optimal investments whenever (36) gives the same technology cutoff $\overline{c}$ as the socially-optimal condition from (5), that is, the price cap $\overline{p}$ and capacity payment $z$ are such that:

$$z - k(\overline{c}) - (\overline{p} - \overline{c})\, k'(\overline{c}) = -(p^* - \overline{c})\, k'(\overline{c}) - k(\overline{c}) - M'(q(\overline{c})),$$

which is the case whenever (9) is satisfied, as claimed.

Next, we wish to establish that the family of socially-optimal instruments $(\overline{p}, z)$ from (9) satisfies the two conditions (32) and (33) in Lemma 2:

*Step 1.* We start with the regularity condition from (32) that the capacity payment should not be too small, which reads:

$$A = z - k(0) - \overline{p}k'(0) > 0.$$

Using (9) to express the capacity payment $z$ in terms of the price cap $\bar{p}$ yields:

$$A = -M'(\hat{\varepsilon}) - (p^* - \bar{p})\, k'(\bar{c}) - k(0) - \bar{p}k'(0)\,.$$

We wish to find a lower bound on $A$, that is, a combination of parameter values for which $A$ is at a minimum—from which it follows that indeed $A > 0$. In particular, we look for the "most critical" $\bar{p}$ for given $\hat{\varepsilon}$ and $\bar{c}$.[20] Under these circumstances, we have that:

$$\frac{dA}{d\bar{p}} = k'(\bar{c}) - k'(0) > 0,$$

as $k''(c) > 0$. Hence, for given $\hat{\varepsilon}$ and $\bar{c}$, $A$ is bounded from below by the case where $\bar{p} \searrow \bar{c}$ and so:

$$\begin{aligned}A &> -M'(\hat{\varepsilon}) - (p^* - \bar{c})\, k'(\bar{c}) - k(0) - \bar{c}k'(0)\\ &= k(\bar{c}) - k(0) - \bar{c}k'(0)\\ &> 0,\end{aligned}$$

[20] As we only need to identify a lower bound, we do not need to consider whether the "worst" combination of parameter is actually consistent with some particular demand distribution.

where the second line follows since we are at a social optimum, (5), and the last line again follows from $k''(c) > 0$. Hence, we conclude that the regularity condition in (32) is satisfied for the socially-optimal instruments from (9).

*Step 2.* Next we consider the regularity condition from (33) that the capacity payment should not be too large. There are two such cases. First, if $p^* > \hat{p} = \bar{p}$, then (33) can be written as:

$$B = z - k(\bar{p}) < 0.$$

We follow a similar approach to Step 1 but now wish to find an upper bound on $B$. Again using (9) to express the capacity payment $z$ in terms of the price cap $\bar{p}$ yields:

$$B = -M'(\hat{\varepsilon}) - (p^* - \bar{p})\, k'(\bar{c}) - k(\bar{p}).$$

For given $\hat{\varepsilon}$ and $\bar{c}$, we have:

$$\frac{dB}{d\bar{p}} = k'(\bar{c}) - k'(\bar{p}) < 0,$$

as $k''(c) > 0$. Hence, for given $\hat{\varepsilon}$ and $\bar{c}$, $B$ is bounded from above by the case where $\bar{p} \searrow \bar{c}$ and so:

$$\begin{aligned} B &< -M'(\hat{\varepsilon}) - (p^* - \bar{c})\, k'(\bar{c}) - k(\bar{c}) \\ &= 0, \end{aligned}$$

where the second line follows from (5). Hence, we conclude that the regularity condition in (33) is satisfied in this case. Second, and finally, if $p^* = \hat{p} < \bar{p}$, then (33) can be written as:

$$D = z - k(p^*) - (\bar{p} - p^*)\, k'(p^*) < 0.$$

Again using (9) to express the capacity payment $z$ in terms of the price cap $\bar{p}$ yields:

$$D = -M'(\hat{\varepsilon}) - (p^* - \bar{p})\, (k'(\bar{c}) - k'(p^*)) - k(p^*).$$

For given $\hat{\varepsilon}$ and $\bar{c}$, we have:

$$\frac{dD}{d\bar{p}} = k'(\bar{c}) - k'(p^*) < 0,$$

as $k''(c) > 0$. Hence, for given $\hat{\varepsilon}$ and $\bar{c}$, $D$ is bounded from above by the case where $\bar{p} \searrow \bar{c}$ and so:

$$\begin{aligned} D &< -M'(\hat{\varepsilon}) - (p^* - \bar{c})\, (k'(\bar{c}) - k'(p^*)) - k(p^*) \\ &= k(\bar{c}) - k(p^*) + (p^* - \bar{c})\, k'(p^*) \\ &< 0, \end{aligned}$$

where the second line follows since we are at a social optimum, (5), and the last line again follows from $k''(c) > 0$. Hence, we conclude that the regularity condition in (33) is satisfied also in this case. Therefore the two regularity conditions in (32) and (33) are both satisfied for the socially-optimal instruments from (9), as required.

**Proof of Proposition 3** Assume that $\bar{p} > p_0^*$ is a notional VOLL level that the social planner uses when optimizing investments. Note that the notional level may differ from the true VOLL level. Hence, we can use the results in Proposition 1 to determine a cutoff $\bar{c}(\bar{p})$ from the condition $-(\bar{p}-\bar{c})\,k'(\bar{c}) - k(\bar{c}) - M'(q(\bar{c})) = 0$ and an associated supply curve $q(c) = F^{-1}(1 + k'(c))$, for $c \in [0, \bar{c}(\bar{p})]$. Next, we can use results in Proposition 2 to establish alternative price caps $\tilde{p}$ and capacity payments that will give the same investments as a social planner would for the notional VOLL level $\bar{p}$. Equation (11) follows from the zero-profit condition in (36).

**Proof of Proposition 4**. For the statement in part (i), implicit differentiation of (11) with respect to the capacity payment $z$ yields:

$$\begin{aligned} 1 - k'(\overline{c})\frac{\partial \overline{c}}{\partial z} + \frac{\partial \overline{c}}{\partial z}k'(\overline{c}) - (\overline{p}-\overline{c})\,k''(\overline{c})\frac{\partial \overline{c}}{\partial z} &= 0 \\ 1 - (\overline{p}-\overline{c})\,k''(\overline{c})\frac{\partial \overline{c}}{\partial z} &= 0 \Longrightarrow \frac{\partial \overline{c}}{\partial z} = \frac{1}{(\overline{p}-\overline{c})\,k''(\overline{c})} \geq 0. \end{aligned}$$

Similarly, implicit differentiation of (36) with respect to the price cap $\overline{c}$ yields:

$$\begin{aligned} -k'(\overline{c})\frac{\partial \overline{c}}{\partial \overline{p}} - \left(1 - \frac{\partial \overline{c}}{\partial \overline{p}}\right)k'(\overline{c}) - (\overline{p}-\overline{c})\,k''(\overline{c})\frac{\partial \overline{c}}{\partial \overline{p}} &= 0 \\ -k'(\overline{c}) - (\overline{p}-\overline{c})\,k''(\overline{c})\frac{\partial \overline{c}}{\partial \overline{p}} &= 0 \Longrightarrow \frac{\partial \overline{c}}{\partial \overline{p}} = \frac{-k'(\overline{c})}{(\overline{p}-\overline{c})\,k''(\overline{c})} \geq 0. \end{aligned}$$

The statement in part (ii) follows straightforwardly using the results from part (i).

**Proof of Proposition 5.** The first-order condition in (10) also holds at $\overline{c}$ and for marginal changes in the capacity $\hat{\varepsilon}$. Hence differentiating with respect to the capacity volume $\hat{\varepsilon}$ gives:

$$-k''(\overline{c})\frac{\partial \overline{c}}{\partial \hat{\varepsilon}} + f(\hat{\varepsilon}) = 0 \Longrightarrow \frac{\partial \overline{c}}{\partial \hat{\varepsilon}} = \frac{f(\hat{\varepsilon})}{k''(\overline{c})} \geq 0.$$

Moreover, implicit differentiation of (11) with respect to $\hat{\varepsilon}$ yields:

$$\begin{aligned} 0 &= \frac{\partial z}{\partial \hat{\varepsilon}} - k'(\overline{c})\frac{\partial \overline{c}}{\partial \hat{\varepsilon}} + \frac{\partial \overline{c}}{\partial \hat{\varepsilon}}k'(\overline{c}) - (\overline{p}-\overline{c})\,k''(\overline{c})\frac{\partial \overline{c}}{\partial \hat{\varepsilon}} \\ &\Longrightarrow \frac{\partial z}{\partial \hat{\varepsilon}} = (\overline{p}-\overline{c})\,k''(\overline{c})\frac{\partial \overline{c}}{\partial \hat{\varepsilon}} = (\overline{p}-\overline{c})\,f(\hat{\varepsilon}) \geq 0. \end{aligned}$$

**Proof of Proposition 6.** The first-order condition in (10) also holds at $\overline{c}$; hence, $\overline{c}$ is fixed if $\hat{\varepsilon}$ is fixed and so $\left.\frac{d\overline{c}}{d\overline{p}}\right|_{\hat{\varepsilon}\text{ fixed}} = 0$. Moreover, implicit differentiation of (11) with respect to $\overline{p}$ shows that $\left.\frac{dz}{d\overline{p}}\right|_{\hat{\varepsilon}\text{ fixed}} = k'(\overline{c}) < 0$.

**Proof of Lemma 1**. Differentiating the condition for optimal investment from (18) shows the impact of more renewables, for any technology level $c$, is given by:

$$-\frac{\partial F(q(c,w);w)}{\partial \varepsilon}\frac{\partial q(c,w)}{\partial w} - \frac{\partial F(q(c,w);w)}{\partial w} = 0,$$

so that:

$$\frac{\partial q(c,w)}{\partial w} = -\frac{\partial F(q(c,w);w)}{\partial w} \Big/ \frac{\partial F(q(c,w);w)}{\partial \varepsilon} < 0. \tag{39}$$

as $\frac{\partial F(q(c,w);w)}{\partial \varepsilon} < 0$ and $\frac{\partial F(q(c,w);w)}{\partial w} > 0$ are assumed.

**Proof of Proposition 7.** Consider first the case where the price cap $\overline{p}$ increases. On one hand, at the margin, this increases the expected revenue to producers from the spot market by $1 - F(\hat{\varepsilon})$, the loss of load probability. On the other hand, this raises the market capacity by $\frac{\partial \widehat{\varepsilon}}{\partial \overline{p}}$ which in turn marginally reduces the loss of load probability and thereby reduces payments to producers below the technology cutoff by $(\overline{p} - \overline{c}) f(\hat{\varepsilon}) \frac{\partial \widehat{\varepsilon}}{\partial \overline{p}}$. We now show that the latter effect exactly offsets the former effect, so that the combined impact on any plant with marginal cost below an initial technology cutoff $\overline{c}_0$ is zero.

It follows from (10) that

$$\frac{\partial \widehat{\varepsilon}}{\partial \overline{c}} = \frac{k''(\overline{c})}{f(\hat{\varepsilon})} \tag{40}$$

and we know from (12) that $\frac{\partial \overline{c}}{\partial \overline{p}} = \frac{-k'(\overline{c})}{(\overline{p}-\overline{c})k''(\overline{c})}$, so we can write the latter effect as:

$$\begin{aligned}(\overline{p} - \overline{c}) f(\hat{\varepsilon}) \frac{\partial \widehat{\varepsilon}}{\partial \overline{p}} &= (\overline{p} - \overline{c}) f(\hat{\varepsilon}) \frac{\partial \overline{c}}{\partial \overline{p}} \frac{\partial \widehat{\varepsilon}}{\partial \overline{c}} \\ &= (\overline{p} - \overline{c}) f(\hat{\varepsilon}) \frac{-k'(\overline{c})}{(\overline{p} - \overline{c}) k''(\overline{c})} \frac{k''(\overline{c})}{f(\hat{\varepsilon})} = -k'(\overline{c}) = 1 - F(\hat{\varepsilon}),\end{aligned}$$

where the last equality uses (10). This shows that the two effects are exactly offsetting.

The argument is similar for the case where the capacity payment $z$ increases. On one hand, at the margin, the capacity payment to production below the technology cutoff $\overline{c}_0$ increases by 1. On the other hand, expected revenues to producers in the spot market decrease by $(\overline{p} - \overline{c}) f(\hat{\varepsilon}) \frac{\partial \widehat{\varepsilon}}{\partial z}$. We again can show that the latter effect exactly offsets the former effect. This here follows directly from (12) and (40).

$$(\overline{p} - \overline{c}) f(\hat{\varepsilon}) \frac{\partial \widehat{\varepsilon}}{\partial z} = (\overline{p} - \overline{c}) f(\hat{\varepsilon}) \frac{\partial \overline{c}}{\partial z} \frac{\partial \widehat{\varepsilon}}{\partial \overline{c}} = \frac{1}{(\overline{p} - \overline{c}) k''(\overline{c})} \frac{k''(\overline{c})}{f(\hat{\varepsilon})} = 1. \tag{41}$$

It follows that, in both cases, any extra revenue due to the higher price cap and/or capacity payments goes solely to covering the production and investment costs of new investments above the initial technology cutoff $\overline{c}_0$.

**Proof of Proposition 8**. Again using (18), conventional supply at the optimal technology cutoff $\overline{c}$ can be written as $q(\overline{c}, w) = F^{-1}(1 + k'(\overline{c}), w)$. Differentiating the condition for the socially-optimal technology cutoff $\overline{c}$ from (19) shows that the impact of $w$ on $\overline{c}$ satisfies:

$$\frac{d\overline{c}}{dw}k'(\overline{c}) - (p^* - \overline{c})\, k''(\overline{c})\frac{d\overline{c}}{dw} - k'(\overline{c})\frac{d\overline{c}}{dw} - \frac{\partial^2 M(q(\overline{c}, w), w)}{\partial \hat{\varepsilon}^2}\frac{\partial q(\overline{c}, w)}{\partial w} - \frac{\partial^2 M(q(\overline{c}, w), w)}{\partial \hat{\varepsilon}\partial w} = 0.$$

This can be rearranged to give:

$$\frac{d\overline{c}}{dw} = -\frac{\frac{\partial^2 M(q(\overline{c},w),w)}{\partial \hat{\varepsilon}^2}\frac{\partial q(\overline{c},w)}{\partial w} + \frac{\partial^2 M(q(\overline{c},w),w)}{\partial \hat{\varepsilon}\partial w}}{(p^* - \overline{c})\, k''(\overline{c})} \geq 0, \tag{42}$$

which is positive given the maintained assumptions $k''(\overline{c}) > 0$ and $\frac{\partial^2 M(q(\overline{c},w),w)}{\partial \hat{\varepsilon}^2} \geq 0$, the renewables assumption $\frac{\partial^2 M(q(\overline{c},w),w)}{\partial \hat{\varepsilon}\partial w} \leq 0$ as well as $\frac{\partial q(\overline{c},w)}{\partial w} < 0$ which holds because (39) holds for all $c$. Again using (18), the socially-optimal loss of loss probability satisfies $LOLP = 1 - F(\hat{\varepsilon}, w) = -k'(\overline{c})$ and so differentiation yields:

$$\frac{dLOLP}{dw} = -\frac{dk'(\overline{c})}{dw} = -k''(\overline{c})\frac{d\overline{c}}{dw} \leq 0.$$

Finally, for the special case with $M(\hat{\varepsilon}, w) \equiv 0$, it follows by inspection that $\frac{d\overline{c}}{dw} = 0$ and so also $\frac{dLOLP}{dw} = 0$.

**Proof of Proposition 9**. The overall impact of renewables on the socially-optimal conventional capacity is given by:

$$\frac{dq(\overline{c}, w)}{dw} = \frac{\partial q(\overline{c}, w)}{\partial \overline{c}}\frac{d\overline{c}}{dw} + \frac{\partial q(\overline{c}, w)}{\partial w}. \tag{43}$$

We derive the first term by differentiating the optimality condition (18) to obtain:

$$-\frac{\partial F(q(c, w), w)}{\partial \varepsilon}\frac{\partial q(c, w)}{\partial c} = -k''(c) \Longrightarrow \frac{\partial q(c, w)}{\partial c} = k''(c) \Big/ \frac{\partial F(q(c, w), w)}{\partial \varepsilon} > 0. \tag{44}$$

We have already derived the second term $\frac{d\overline{c}}{dw}$ in (42) and the third term $\frac{\partial q(\overline{c},w)}{\partial w}$ in (39) so using these gives:

$$\begin{aligned}
\frac{dq(\overline{c}, w)}{dw} &= \frac{\frac{\partial^2 M(q(\overline{c},w),w)}{\partial \hat{\varepsilon}^2}\frac{\frac{\partial F(q(c,w);w)}{\partial w}}{\frac{\partial F(q(c,w);w)}{\partial \varepsilon}} - \frac{\partial^2 M(q(\overline{c},w),w)}{\partial \hat{\varepsilon}\partial w}}{\frac{\partial F(q(c,w),w)}{\partial \varepsilon}(p^* - \overline{c})} - \frac{\frac{\partial F(q(c,w);w)}{\partial w}}{\frac{\partial F(q(c,w);w)}{\partial \varepsilon}} \\
&= -\frac{1}{\frac{\partial F(q(c,w),w)}{\partial \varepsilon}(p^* - \overline{c})}\left(\frac{\partial^2 M(\hat{\varepsilon}, w)}{\partial \hat{\varepsilon}\partial w} + \frac{\partial F(q(c, w); w)}{\partial w}\left[(p^* - \overline{c}) - \frac{\frac{\partial^2 M(q(\overline{c},w),w)}{\partial \hat{\varepsilon}^2}}{\frac{\partial F(q(c,w),w)}{\partial \varepsilon}}\right]\right) \\
&\leq 0,
\end{aligned}$$

which holds if and only if $\left[(p^* - \bar{c}) - \frac{\partial^2 M(\hat{\varepsilon},w)}{\partial \hat{\varepsilon}^2} \Big/ \frac{\partial F(\hat{\varepsilon},w)}{\partial \varepsilon}\right] \frac{\partial F(\hat{\varepsilon},w)}{\partial w} \geq -\frac{\partial^2 M(\hat{\varepsilon},w)}{\partial \hat{\varepsilon} \partial w}$, as claimed.

**Proof of Proposition 10**. The family $(\overline{p}, z)$ of socially-optimal policies is defined by the condition in (20) so, for a fixed $\overline{p}$, we have:

$$\frac{\partial z(w,\overline{p})}{\partial w} + \frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon}^2} \frac{\partial q(\bar{c},w)}{\partial w} + \frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon} \partial w} + (p^* - \overline{p}) k''(\bar{c}) \frac{d\bar{c}}{dw} = 0,$$

and so:

$$\frac{\partial z(w,\overline{p})}{\partial w} = -\frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon}^2} \frac{\partial q(\bar{c},w)}{\partial w} - \frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon} \partial w} - (p^* - \overline{p}) k''(\bar{c}) \frac{d\bar{c}}{dw}.$$

Using the term for $\frac{d\bar{c}}{dw}$ from (42) we obtain:

$$\begin{aligned}
\frac{\partial z(w,\overline{p})}{\partial w} &= -\frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon}^2} \frac{\partial q(\bar{c},w)}{\partial w} - \frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon} \partial w} \\
&\quad + (p^* - \overline{p}) \frac{\frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon}^2} \frac{\partial q(\bar{c},w)}{\partial w} + \frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon} \partial w}}{(p^* - \bar{c})} \\
&= -\left(\frac{\overline{p} - \bar{c}}{p^* - \bar{c}}\right) \left[\frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon}^2} \frac{\partial q(\bar{c},w)}{\partial w} + \frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon} \partial w}\right] \geq 0,
\end{aligned}$$

given the assumptions $\frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon}^2} \geq 0$ and $\frac{\partial^2 M(q(\bar{c},w),w)}{\partial \hat{\varepsilon} \partial w} \leq 0$ together with $\frac{\partial q(c,w)}{\partial w} < 0$ from (39). Finally, for the special case with $M(\hat{\varepsilon}, w) \equiv 0$, it follows by inspection that $\frac{\partial z(w,\overline{p})}{\partial w} = 0$.

**Proof of Proposition 11.** Under the design laid out in Section 6.1., the non-reserve market is independent from the reserve market. Hence, revenues for plants in the energy-only market are identical to those in a competitive market with the same price cap $\overline{p}$ and no capacity payment.[21] Hence, the invested capacity $\widehat{\varepsilon}$, technology cutoff $\bar{c}$ and technology mix will also be the same. When solving for investments into the reserve, we can take investments into the non-reserve market as given. We know from Proposition 7 that even if a capacity payment $z$ was introduced, this would not change the technology mix and supply below the marginal cost $\bar{c}$. It follows that revenues for reserve plant are identical to those for plant in the capacity range $[\widehat{\varepsilon}, \widehat{\varepsilon}_r]$ of a US-style capacity market with a uniform capacity payment $z$. Hence, the technology mix for that range and the technology cutoff $\bar{c}_r$ follow from our results for US-style capacity markets. Therefore, in

[21]Recall that $p_0^*$ is the lowest VOLL level for which we can ensure positive socially optimal investments. Hence, $\bar{c}(p_0^*)$ and $q(\bar{c}(p_0^*))$ are the technology cutoff and market capacity, respectively, for that lowest VOLL level. It follows from Proposition 2 that those investments will also occur for $\overline{p} = p_0^* + M'(q(\bar{c}(p_0^*)))/k'(\bar{c}(p_0^*))$ when $z = 0$. We need $\overline{p} > p_0^* + M'(q(\bar{c}(p_0^*)))/k'(\bar{c}(p_0^*))$ to make sure that investments are non-negative.

sum, statements (i)-(iv) follow from Proposition 2. Finally, we verify using comparative statics the statement (v) that the payment difference $\overline{p} - p_r$ (which energy-only plant make relative to reserve plant, in situations when the strategic reserve is used) is equal in expectation to the capacity payment $z$ to reserve plant. By the arguments of (41), a marginally higher capacity payment to reserve plant would lower their spot-market revenues by $(\overline{p} - \overline{c}) f(\hat{\varepsilon}) \frac{\partial \widehat{\varepsilon}}{\partial z} = 1$, so that expected profit remains zero. The spot price for energy-only plants does not change when the capacity payment $z$ to reserve plants increases. Hence, the payment difference $\overline{p} - p_r$ will, in expectation, give an extra payment to non-reserve plant, relative to reserve plant, which will increase at the same rate as $z$ increases.