WORKING PAPERS

N° IDEI-815

January 2014

The "demand side" effect of price caps: uncertainty, imperfect competition, and rationing

Thomas-Olivier Léautier
Toulouse School of Economics (IAE, IDEI, CRM)

INSTITUT D'ECONOMIE INDUSTRIELLE

Toulouse School of Economics

The "demand side" effect of price caps: uncertainty, imperfect competition, and rationing

Thomas-Olivier Léautier
Toulouse School of Economics (IAE, IDEI, CRM)
21 allée de Brienne
F31 000 Toulouse
thomas.leautier@tse-fr.eu

January 27, 2014

Abstract

Price caps are often used by policy makers to "regulate markets". Previous analyses have focussed on the "supply side" impact of these caps, and derived the optimal price cap, which maximizes investment and welfare. This article expands the analysis to include the "demand side" impact of price caps: when prices can no longer rise, customers must be rationed to adjust demand to available supply. This yields two new findings, that contradict previous analyses. First, the welfare-maximizing cap is higher than the capacity-maximizing cap, since increasing the cap increases gross surplus when customers are rationed. Second, in somes cases, the capacity-maximizing cap leads to lower capacity and welfare than no cap. These findings underscores the importance for policy makers to examine the impact on customers when they impose price caps.

Keywords: price caps, imperfect competition, rationing, investment incentives

JEL Classification: L13, L94

1 Introduction

Price caps are widely used policy instruments. In electricity markets for example, System Operators and policy makers impose caps on prices, either formally or through operating practices (Joskow (2007)). Rent controls, in place in various jurisdictions, constitute a form of price cap. Regulation of infrastructure, for example telecommunication and electric power networks, often relies on price caps. Prices of medical procedures and drugs are capped in many countries (Schut and Van de Ven (2005), Mougeot and Naegelen (2005)). Salary caps exist in many sport leagues (Késenne (2000)). Following the 2008 financial crisis, the possibility of capping extremely high compensations, for example those of traders and *CEOs*, has been discussed in Europe and in the United States. Economic analysis of price caps (reviewed below) has sofar focussed on their impact on supply. By extending the analysis to include their impact on demand, in particular rationing, this article provides new insights. Policy makers should therefore carefully assess the demand-side impact of price caps in decision making.

Economists have long held reservations about the use of price caps. The argument can be framed using a simple two-period model[1]. Firms (or individuals) invest in period 1, and receive profits in period 2, during which a cap is imposed on prices. Conventional wisdom is that increasing the cap increases prices, hence profits in period 2, hence investment in period 1, hence lead to higher overall welfare.

While this result holds if the industry considered is perfectly competitive, matters are more complex if competition is imperfect and demand in period 2 uncertain. Earle et al. (2007) and Zöttl (2011) show that, under mild assumptions on the shape of the demand function, contrary to the previous result, increasing the cap in a certain range *reduces* investment.

The intuition for this surprising result is that price caps also have a second impact on investment incentives: increasing the cap reduces the probability that the cap is binding, which reduces expected profits, hence investment. Earle et al. (2007) show that the first effect dominates if the price cap is close the marginal cost, while the second effect dominates if the cap is sufficiently high. Building on this analysis, Zöttl (2001) derives the capacity-maximizing price cap for a certain class of demand functions, and shows that this cap leads to higher investment and welfare than the absence of cap. These results are extremely important, as they provide strong justification for the use of price caps.

However, the previous analysis is incomplete, that ignores the demand-side impact of price caps: when a cap is imposed, demand must be rationed. By including rationing in the analysis, this article shows that price caps are less beneficial than suggested by the previous literature.

[1] Recent articles, for example Evans and Guthrie (2012), and Roques and Favva (2009), have examined the dynamic effects of price caps. While they refine our understanding of the supply-side effects of price caps, they have not addressed the demand side effects, which is the contribution of this article.

Consider, as in Earle et al. (2007) and Zöttl (2011) that demand in period 2 varies, and is uncertain at the time of investment in period 1. When the cap is binding, prices can no longer rise, hence cannot reduce demand. Some form of rationing is required to adjust demand to available supply. In power markets, System Operators administratively decide which customers to cut, a practice known as rolling blackouts. In other markets, first-come first-served is the method of rationing. In other situations, non monetary benefits, such as knowledge of the sellers, family ties, affiliation to a political party, or social status may be used for rationing.

The marginal welfare impact of increasing the price cap is thus the sum of two terms: (*i*) the marginal impact on capacity, as identified by Earle et al. (2007) and Zöttl (2011), and (*ii*) the marginal impact on gross consumers surplus. The capacity maximizing price cap sets first the term to zero. Under reasonable assumptions, increasing the cap increases gross surplus when customers are rationed: demand is lower, hence less stringent rationing is required. Thus, welfare increases by increasing the cap from the capacity maximizing level. This intuition is formalized in Proposition 1.

The analysis can be further expanded. Suppose that a fraction of consumers does not respond to prices. This is the case in the power industry, where some customers face a fixed retail price, that does not vary in real-time to follow the wholesale price. This is also the case in the health-care industry, where the price for most drugs and procedures does not vary as demand varies. Demand is adjusted to available supply through two channels: price increase for price-reactive customers, and rationing for constant-price customers. The need to maintain consistency between these two adjustment mechanisms imposes an upper limit on admissible price caps.

In this setting, Proposition 2 derives a sufficient condition for the capacity-maximizing cap to induce lower capacity than no cap, even under Zöttl (2011) sufficient conditions.

Finally, by combining Propositions 1 and 2, Proposition 3 proves that, if the capacity-maximizing cap leads to lower capacity than no cap, imposition of a price cap reduces welfare. Thus, by including their impact on rationing Proposition 3, weakens previous results on the desirability of price caps.

The situation described in Proposition 3 is empirically relevant, as illustrated on a simplified representation of the French electric power market.

The policy implications of this analysis are straightforward: when considering price caps, policy makers and analysts should not only examine their impact on supply, but also their impact on demand, in particular rationing of excess demand. The welfare impact of this rationing should be evaluated, as it may exceed the supply-side benefits of the cap.

This article brings together two strands of the literature. On the supply side, the analysis of price cap under uncertainty was initially developed by Earle et al. (2007). The peak-load pricing analysis, that underpins the analysis, was initially developed by Boiteux (1949), and Crew and Kleindorfer (1976). This article builds on the previously discussed two-stage Cournot model developed by Zöttl (2011), which incorporate price caps into peak-load pricing.

On the demand side, this article applies the formalism of rationing, in particular the Value of Lost Load, developed in the electricity economics literature, as summarized for example by Joskow and Tirole (2007) and Stoft (2002). The dichotomy between "price reactive" customers and "constant price customers" was developed by Borenstein and Holland (2005) and Joskow and Tirole (2007).

This article is structured as follows. Section 2 presents the setup and the equilibrium Cournot investment. It expends Zöttl (2011) by considering an additional – and empirically relevant – case. Section 3 proves the difference between the capacity-maximizing and the welfare-maximizing caps. Section 4 extends the analysis to include constant price customers. Section 5 proves that, in this case, no price cap may sometimes yields higher welfare than the capacity maximizing cap. Section 6 illustrates the analysis on the French power markets. Finally, Section 7 presents concluding remarks and avenues for further research. Technical proofs are presented in the Appendix.

2 Cournot competition with uncertain demand and a price cap

2.1 Demand, supply, and curtailment

Underlying demand All customers are homogenous. Individual demand is $D(p,t)$, where p is the electricity price, and $t \geq 0$ is the state of the world, distributed according to cumulative distribution $F(.)$, and probability distribution $f(.) = F'(.)$, common to all stakeholders.

Inverse demand $P(Q;t)$ is defined by $D(P(q;t);t) = Q$. Gross consumers surplus is $S(p;t) = \int_0^{D(p;t)} P(q;t)\,dq$.

Assumption 1 *$\forall t \geq 0$, $\forall q \leq Q$, the inverse demand $P(Q;t)$ satisfies*

$$P_q(Q,t) < \min(0, -qP_{qq}(Q,t)) \text{ and } P_t(Q,t) > q|P_{qt}(Q,t)| \geq 0.$$

$P_q < 0$ requires inverse demand to be downward sloping. $P_q(Q,t) < -qP_{qq}(Q,t)$ implies that the marginal revenue is decreasing with output

$$\frac{\partial^2}{\partial q^2}(qP(Q,t)) = 2P_q(Q,t) + qP_{qq}(Q,t) < 0,$$

and guarantees existence and unicity of a Cournot equilibrium.

$P_t > 0$ orders the states of the world, $P_t(Q,t) > q|P_{qt}(Q,t)|$ implies that the marginal revenue is increasing with the state of the world

$$\frac{\partial^2}{\partial t \partial q}(qP(Q,t)) = P_t(Q,t) + qP_{qt}(Q,t) > 0,$$

and that the Cournot output and profit (defined later) are increasing.

Assumption 1 is met for example if demand is linear with constant slope $P(Q,t) = a(t) - bQ$, with $b > 0$ and $a'(t) > 0$.

In this and the next Section, all customers face and respond to the spot price. In Sections 4 and later, we introduce constant price customers.

Supply This article considers a single production technology, with marginal cost $c > 0$ and investment cost r. The peak load pricing literature proves that a single technology is sufficient to analyze total installed capacity, that depends solely on the characteristics of the marginal technology (see for example Boiteux (1949) for the perfect competition case and Zöttl (2011) for the imperfect competition case).

I assume consuming the first unit is valuable $P(0,t) > c$ and $\mathbb{E}[P(0,t)] > (c+r)$. This guarantees existence of a positive equilibrium investment.

Imperfect competition The industry is composed of N producers, that play a two-stage game: in stage 1, producer n installs capacity k^n; in stage 2 he produces $q^n(t) \leq k^n$ in state t and sells it entirely in the spot market. Producers are assumed to compete à la Cournot in the spot markets, facing inverse demand $P(Q,t)$. Stage 2 can be interpreted as a repetition of multiple states of the world over a given period (for example one year), drawn from the distribution $F(.)$.

The game is solved by backwards induction: producers first compute spot market profits from a Nash equilibrium given installed capacities $\left(k^1,...,k^N\right)$ in each state of the world t; then they make their investment choice in stage 1 using the expectation of these spot market profits.

$Q(t) = \sum_{n=1}^{N} q^n(t)$ and $K = \sum_{n=1}^{N} k^n$ are respectively aggregate production in state t and aggregate installed capacity. K^C is the equilibrium capacity referred to as the "Cournot" capacity in the following.

Price cap and curtailment To limit the exercise of market power, policy makers impose a cap $\bar{p}^W$ on the spot price. In power markets, the price cap may be a formal cap, or the result of operational practices that depress prices (see Joskow (2007)). The price cap must be higher than the long-term marginal cost of the first unit, otherwise it would block any investment. At the other extreme, the price cap must be binding in some states of the world, otherwise it would be useless. Anticipating on Lemma 1 below, define $K^C(p)$ the Cournot capacity when the price cap is p. The highest binding cap is p^∞, uniquely defined by

$$\lim_{t\to+\infty} P\left(K^C(p^\infty),t\right) = p^\infty.$$

Then, $\bar{p}^W$ verifies

$$(c+r) \leq \bar{p}^W \leq p^\infty.$$

The first state of the world for which the price cap is binding for production Q is denoted $\hat{t}_0\left(Q,\bar{p}^W\right)$, and uniquely defined by

$$P\left(Q;\hat{t}_0\left(Q,\bar{p}^W\right)\right) = \bar{p}^W.$$

If the price cap is never binding, $\hat{t}_0\left(Q,\bar{p}^W\right) \to +\infty$.

For states of the world where the cap is binding, the price cannot rise to lower demand. When this occurs and the firms produce at capacity, administrative action is required to curtail (or ration) a fraction of customers, hence to reduce demand to match available supply. To describe the economics or rationing, I use the formalism developed in the electric power literature, that can be applied to other industries.

The *serving ratio* is denoted $\gamma \in [0,1]$: $\gamma = 0$ means full curtailment, while $\gamma = 1$ means no curtailment. For state t, $\mathcal{D}(p,\gamma,t)$ is the demand for (fixed) price p and serving ratio γ, and $\mathcal{S}(p,\gamma,t)$ is the gross consumer surplus. By construction, $\mathcal{D}(p,1,t) \equiv D(p,t)$ and $\mathcal{S}(p,1,t) \equiv S(p,t)$.

Since customers are homogeneous, there is no basis to discriminate among them. Thus, curtailment is proportional, for example proceeds by geographic zones. This yields

$$\mathcal{D}(p,\gamma,t) = \gamma D(p,t).$$

Joskow and Tirole (2007) illustrate on a simple example how the net surplus $(\mathcal{S}(p,\gamma;t) - p\mathcal{D}(p,\gamma,t))$ depends on the information consumers hold about the curtailment. Suppose first rationing is perfectly anticipated, i.e., consumers know exactly who will be rationed. The fraction γ of customers that will not be curtailed consumes normally, hence receives surplus $S(p,t)$ per customer. The fraction $(1-\gamma)$ of customers that will be curtailed does not attempt to consume, hence receives no surplus. At the aggregate level, this yields

$$\mathcal{S}(p,\gamma;t) = \gamma S(p;t).$$

If rationing is not perfectly anticipated, customers who end up not being curtailed may refrain from consuming, hence receive no surplus. Conversely, consumers that end up being curtailed may attempt to consume, hence derive a negative surplus when they are curtailed (e.g., they step in an elevator and find out power is cut).

When curtailment occurs, the Value of Lost Load ($VoLL$) represents the value consumers would place on an extra unit. Formally, it is defined as

$$v(p,\gamma,t) = \frac{\frac{\partial \mathcal{S}}{\partial \gamma}}{\frac{\partial \mathcal{D}}{\partial \gamma}}(p,\gamma,t).$$

If rationing is perfectly anticipated, the $VoLL$ is always higher than the price:

$$v(p,\gamma,t) = \frac{\frac{\partial \mathcal{S}}{\partial \gamma}}{\frac{\partial \mathcal{D}}{\partial \gamma}}(p,\gamma,t) = \frac{S(p,t)}{D(p,t)} > p,$$

and is increasing in price:

$$\frac{\partial v}{\partial p} = \frac{D(p,t)\frac{\partial S(p,t)}{\partial p} - S(p,t)\frac{\partial D(p,t)}{\partial p}}{(D(p,t))^2} = \frac{pD(p,t) - S(p,t)}{(D(p,t))^2}\frac{\partial D(p,t)}{\partial p} = \frac{p - v(p,\gamma,t)}{D(p,t)}\frac{\partial D(p,t)}{\partial p} > 0.$$

It seems reasonable to assume these properties also hold when rationing is not perfectly anticipated. Consumers are always willing to pay at least as much for one unit when they are curtailed as they are when they are not curtailed. *Ceteris paribus*, this value increases as price increases.

When curtailment occurs, the serving ratio for installed capacity K and constant price p is determined such that

$$\mathcal{D}(p,\gamma^*,t) = K.$$

The impact of increasing p on consumers' surplus is

$$\frac{d\mathcal{S}}{dp} = \frac{\partial \mathcal{S}}{\partial p} + \frac{\partial \mathcal{S}}{\partial \gamma}\frac{\partial \gamma^*}{\partial p} = \frac{\partial \mathcal{S}}{\partial p} - v(p,\gamma^*,t)\frac{\partial \mathcal{D}}{\partial p},$$

since, from the definition of γ^*,

$$\frac{\partial \gamma^*}{\partial p} = -\frac{\frac{\partial \mathcal{D}}{\partial p}}{\frac{\partial \mathcal{D}}{\partial \gamma}} \Rightarrow \frac{\partial \mathcal{S}}{\partial \gamma}\frac{\partial \gamma^*}{\partial p} = -\frac{\partial \mathcal{S}}{\partial \gamma}\frac{\frac{\partial \mathcal{D}}{\partial p}}{\frac{\partial \mathcal{D}}{\partial \gamma}} = -v(p,\gamma^*,t)\frac{\partial \mathcal{D}}{\partial p}.$$

An increase in price p reduces the gross surplus, but also reduces demand, valued at v. If rationing is anticipated,

$$\frac{\partial \mathcal{S}}{\partial p} - v\frac{\partial \mathcal{D}}{\partial p} = \gamma p\frac{\partial D}{\partial p} - \gamma v\frac{\partial D}{\partial p} = -\gamma\frac{\partial D}{\partial p}(v-p) > 0,$$

the demand effect is larger than the surplus effect, hence an increase in price increases the gross surplus. Again, it seems reasonable that this assumption holds for other rationing technologies. These points are formalized in the following:

Assumption 2 *The $VoLL$ is higher than the price, and,* ceteris paribus*, increasing in price:*

$$v(p,\gamma,t) > p \text{ and } \frac{\partial v}{\partial p} > 0.$$

The total impact of an increase in price is an increase in gross consumer surplus:

$$\frac{d\mathcal{S}}{dp} = \frac{\partial \mathcal{S}}{\partial p} + \frac{\partial \mathcal{S}}{\partial \gamma}\frac{\partial \gamma^*}{\partial p} > 0.$$

As previously mentioned, the formalism of $VoLL$ has been developed in the power industry (hence its name). It applies more generally in any situation where rationing occurs.

2.2 Equilibrium investment

Suppose first no price cap is imposed. The peak load pricing literature shows that two periods must be separated in describing the structure of the equilibrium: off-peak, when production is lower than capacity, firms play a standard Cournot equilibrium with marginal cost c. Aggregate production in state t is the Cournot output, denoted $Q^C(c,t)$, and defined by

$$P\left(Q^C, t\right) + \frac{Q^C}{N} P_q\left(Q^C, t\right) = c.$$

Wholesale price is $P\left(Q^C(c,t), t\right)$, and individual profit is $\frac{Q^C(c,t)}{N}\left(P\left(Q^C(c,t), t\right) - c\right)$.

The first on-peak state of the world, denoted $\hat{t}(K,c)$, is such that the Cournot output is equal to the aggregate capacity K:

$$Q^C\left(c,\hat{t}\right) = K \Leftrightarrow P\left(Q^C\left(c,\hat{t}\right), \hat{t}\right) + \frac{Q^C}{N}\left(c,\hat{t}\right) P_q\left(Q^C\left(c,\hat{t}\right), \hat{t}\right) = c.$$

$\hat{t}(K,c)$ is increasing in both arguments by inspection.

On-peak, in state $t \geq \hat{t}(K,c)$, firms produce at capacity, wholesale price is $P(K,t)$, and individual profit is $\frac{K}{N}(P(K,t) - c)$.

Suppose now a price cap is imposed. Two situations may occur: either the price cap is binding off-peak, i.e., the Cournot price reaches the cap before the Cournot output reaches aggregate capacity, or the price cap is binding on-peak, i.e., the Cournot output reaches aggregate capacity before the price reaches the cap. The latter situation occurs if and only if

$$\hat{t}(K,c) \leq \hat{t}_0\left(K, \bar{p}^W\right). \tag{1}$$

It is more likely is competition is stronger (N larger) and the cap higher. For example, as $N \to +\infty$, $\hat{t}(.,.) \to \hat{t}_0(.,.)$, and condition (1) is always met since $c < \bar{p}^W$.

The marginal value of capacity when all producers invest $k^n = \frac{K}{N}$ is denoted $\Psi\left(K, \bar{p}^W\right)$. The functional form of $\Psi\left(K, \bar{p}^W\right)$ depends on which constraint is binding first. $\Psi\left(K, \bar{p}^W\right)$ is thus defined piecewise, as shown below:

Lemma 1 *The equilibrium capacity $K^C\left(\bar{p}^W\right)$ is characterized by:*

$$\Psi\left(K^C, \bar{p}^W\right) = 0$$

where $\Psi\left(K, \bar{p}^W\right)$ is defined as follows:

1. *If the price cap is reached on-peak (i.e., $\hat{t}(K,c) \leq \hat{t}_0\left(K, \bar{p}^W\right)$),*

$$\Psi\left(K, \bar{p}^W\right) = \Psi_1\left(K, \bar{p}^W\right) = \int_{\hat{t}(K,c)}^{\hat{t}_0\left(K,\bar{p}^W\right)} \left(P(K;t) + \frac{K}{N} P_q(K;t) - c\right) f(t)\, dt + \int_{\hat{t}_0(K,\bar{p}^W)}^{+\infty} \left(\bar{p}^W - c\right) f(t)\, dt - r. \tag{2}$$

2. *If the price cap is reached off-peak (i.e., $\hat{t}_0\left(K, \bar{p}^W\right) < \hat{t}(K,c)$),*

$$\Psi\left(K, \bar{p}^W\right) = \Psi_2\left(K, \bar{p}^W\right) = \int_{\hat{t}_0(K,\bar{p}^W)}^{+\infty} \left(\bar{p}^W - c\right) f(t)\, dt - r. \tag{3}$$

Proof. *If the price cap is reached on-peak, Zöttl (2011)derives $\Psi_1\left(K, \bar{p}^W\right)$ and proves the solution of $\Psi_1\left(K, \bar{p}^W\right) = 0$ is the unique symmetric equilibrium. Intuition for $\Psi_1\left(K, \bar{p}^W\right)$ is as follows. A marginal capacity increase has no impact off-peak. On peak, it generates marginal profit $(p - c)$. If the cap is not binding, it also yields a price reduction, hence the net effect is $\left(P(K;t) + \frac{K}{N} P_q(K;t) - c\right)$. When the cap is binding, there is price no reduction, and the net effect is $\left(\bar{p}^W - c\right)$.*

The formal proof for the price cap reached off-peak is presented in Appendix A. Intuition for $\Psi_2\left(K, \bar{p}^W\right)$ is as follows. When the cap is binding, the wholesale price is set at the cap. Off-peak, firms produce below capacity, hence a marginal increase in capacity has no impact of profits. On-peak arises when demand for price $\bar{p}^W$ reaches capacity, i.e., for $t = \hat{t}_0\left(K, \bar{p}^W\right)$. On-peak, firms produce at capacity. A marginal increase in capacity generates additional margin $\left(\bar{p}^W - c\right)$. This yields $\Psi_2\left(K, \bar{p}^W\right)$. Appendix A proves this constitutes the unique symmetric equilibrium. ■

Lemma 1 extends Zöttl (2011) Theorem 1 by including the case $\hat{t}_0\left(K, \bar{p}^W\right) < \hat{t}(K, c)$. This is an empirically relevant extension: since power markets are highly concentrated, and price caps often very low, imperfectly competitive price may reach the cap before generation produces at capacity. This occurs for example in the numerical illustration presented in Section 6.

One surprising implication of Lemma 1 is that, if the cap is reached off-peak, the optimal capacity does not depend on the number of firms. This is counter-intuitive for Cournot games. The intuition is that, since the cap binding off-peak, on-peak arises when demand at the price cap equals the capacity (i.e., $\hat{t}_0\left(K, \bar{p}^W\right)$), which does not depend on the number of firms. Then, on-peak, price is set at the cap, which again does not depend on the number of firms.

This result underscores the complexity of the analysis of the imposition of price caps.

3 Capacity and welfare-maximizing caps for price reactive consumers

3.1 Capacity-maximizing cap

Differentiation of first-order conditions (2) and (3) yield:

$$\frac{dK^C}{dp} = \frac{1 - F\left(\hat{t}_0\left(K^C(p), p\right)\right) - (p-c) f\left(\hat{t}_0\left(K^C(p), p\right)\right) \frac{\partial \hat{t}_0\left(K^C(p), p\right)}{\partial p}}{\left(-\frac{\partial \Psi}{\partial K}\right)}. \tag{4}$$

where $\frac{\partial \Psi}{\partial K} = \frac{\partial \Psi_1}{\partial K}$ (resp. $\frac{\partial \Psi}{\partial K} = \frac{\partial \Psi_2}{\partial K}$) if the price cap is reached on-peak (resp. off-peak). The denominator is positive. The first term of the numerator is positive and the second term negative. These two terms illustrate the two impacts of a price cap on installed capacity: on the one hand, increasing the cap raises the per unit profit when the cap is binding, hence increases investment incentives. On the other hand, it reduces the probability that the cap is binding, hence the firm loses the margin $(p-c)$ for the relevant states of the world. As observed by Earle and al. (2007), under reasonable assumptions on demand and distribution of states of the world, the first effect dominates when the cap is close to $(c+r)$, while the second effect dominates when the price is high enough. This produces a capacity-maximizing cap, as derived by Zöttl (2011). For the reader's convenience, the proof is presented in the Appendix B. Its main steps are summarized below.

$K^C(p)$ is continuously differentiable on the compact set $[(c+r), p^\infty]$, thus, admits a maximum. If $P_t(D((c+r), 0), 0) > rf(0)$, which I assume holds, $\frac{dK^C}{dp}(c+r) > 0$, hence $(c+r)$ is not a maximum.

p^{∞} is not a maximum either: we can find $\bar{p}^W < p^{\infty}$ such that $K^C(p^{\infty}) < K^C(\bar{p}^W)$. Therefore, the maximum of $K^C(p)$ is interior, hence verifies $\frac{dK^C}{dp} = 0$. This constitutes a necessary condition, as $K^C(p)$ may admit multiple extrema.

3.2 Welfare-maximizing cap

The social welfare depends on which constraint is reached first. Suppose first the cap is reached on-peak. Social welfare is then

$$\begin{aligned} W(K,p) &= W_1(K,p) = \int_0^{\hat{t}(K,c)} \left(S\left(P\left(Q^C,t\right),t\right) - cQ^C\right) f(t)\,dt + \int_{\hat{t}(K,c)}^{\hat{t}_0\left(K,\bar{p}^W\right)} \left(S\left(P\left(K,t\right),t\right) - cK\right) f(t)\,dt \\ &\quad + \int_{\hat{t}_0(K,\bar{p}^W)}^{+\infty} \left(\mathcal{S}\left(\bar{p}^W,\gamma^*,t\right) - cK\right) f(t)\,dt - rK, \end{aligned}$$

where the serving ratio γ^* is such that $\mathcal{D}\left(\bar{p}^W,\gamma^*,t\right) = K$. The first term corresponds to off-peak, the second to on-peak before the cap is binding, and the last to on-peak when the cap is binding and rationing occurs, in which case, consumers surplus is $\mathcal{S}\left(\bar{p}^W,\gamma^*,t\right)$. Since output is a continuous function of the state of the world, so is surplus. Only the derivatives of the integrands appear in $\frac{dW}{d\bar{p}^W}$. Thus,

$$\begin{aligned} \frac{dW_1}{d\bar{p}^W} &= \frac{\partial W_1}{\partial K}\frac{dK^C}{d\bar{p}^W} + \frac{\partial W_1}{\partial \bar{p}^W} \\ &= \left(\int_{\hat{t}(K^C(\bar{p}^W),c)}^{\hat{t}_0\left(K^C(\bar{p}^W),\bar{p}^W\right)} \left(P\left(K^C\left(\bar{p}^W\right),t\right) - c\right) f(t)\,dt + \int_{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)}^{+\infty} \left(\frac{\partial \mathcal{S}}{\partial \gamma}\frac{\partial \gamma^*}{\partial K} - c\right) f(t)\,dt - r\right) \frac{dK^C}{d\bar{p}^W} \\ &\quad + \int_{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)}^{+\infty} \left(\frac{\partial \mathcal{S}}{\partial p} + \frac{\partial \mathcal{S}}{\partial \gamma}\frac{\partial \gamma^*}{\partial p}\right) f(t)\,dt. \end{aligned}$$

Aggregate capacity has no impact on the off-peak surplus, hence only on-peak terms appears in $\frac{dW_1}{d\bar{p}^W}$. Expression of the last integrand has already been computed. To compute the second integrand, observe that, from the definition of γ^*,

$$\frac{\partial \gamma^*}{\partial K} = \frac{1}{\frac{\partial \mathcal{D}}{\partial \gamma}} \Rightarrow \frac{\partial \mathcal{S}}{\partial \gamma}\frac{\partial \gamma^*}{\partial K} = \frac{\frac{\partial \mathcal{S}}{\partial \gamma}}{\frac{\partial \mathcal{D}}{\partial \gamma}} = v\left(p,\gamma^*,t\right).$$

Thus,

$$\frac{dW_1}{d\bar{p}^W} = \left(\int_{\hat{t}(K^C(\bar{p}^W),c)}^{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)} \left(P\left(K^C\left(\bar{p}^W\right),t\right)-c\right) f(t)\,dt + \int_{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)}^{+\infty} \left(v\left(\bar{p}^W,\gamma^*,t\right)-c\right) f(t)\,dt - r\right) \frac{dK^C}{d\bar{p}^W}$$
$$+ \int_{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)}^{+\infty} \left(\frac{\partial \mathcal{S}}{\partial p} - v\left(\bar{p}^W,\gamma^*,t\right)\frac{\partial \mathcal{D}}{\partial p}\right) f(t)\,dt.$$

Observing that $v\left(\bar{p}^W,\gamma^*,t\right)-c = \left(v\left(\bar{p}^W,\gamma^*,t\right)-\bar{p}^W\right)+\left(\bar{p}^W-c\right)$ and inserting the first-order condition defining K^C yields

$$\frac{dW_1}{d\bar{p}^W} = \left(\int_{\hat{t}(K^C(\bar{p}^W),c)}^{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)} -\frac{K^C}{N} P_q\left(K^C,t\right) f(t)\,dt + \int_{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)}^{+\infty} \left(v\left(\bar{p}^W,\gamma^*,t\right)-\bar{p}^W\right) f(t)\,dt\right) \frac{dK^C}{d\bar{p}^W} \quad (5)$$
$$+ \int_{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)}^{+\infty} \left(\frac{\partial \mathcal{S}}{\partial p} - v\left(\bar{p}^W,\gamma^*,t\right)\frac{\partial \mathcal{D}}{\partial p}\right) f(t)\,dt.$$

The term multiplying $\frac{dK^C}{d\bar{p}^W}$ is positive, since $v\left(\bar{p}^W,\gamma^*,t\right) > \bar{p}^W$ by Assumption 2. The last integral is positive by Assumption 2.

A similar analysis shows that, if the price cap is reached off-peak,

$$W(K,p) = W_2(K,p) = \int_0^{\hat{t}_0(Q^C,\bar{p}^W)} \left(S\left(P\left(Q^C,t\right),t\right)-cQ^C\right) f(t)\,dt$$
$$+ \int_{\hat{t}_0(Q^C,\bar{p}^W)}^{\hat{t}_0(K,\bar{p}^W)} \left(S\left(\bar{p}^W,t\right)-cD\left(\bar{p}^W,t\right)\right) f(t)\,dt + \int_{\hat{t}_0(K,\bar{p}^W)}^{+\infty} \left(\mathcal{S}\left(\bar{p}^W,\gamma^*,t\right)-cK\right) f(t)\,dt - rK,$$

and

$$\frac{dW_2}{d\bar{p}^W} = \left(\int_{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)}^{+\infty} \left(v\left(\bar{p}^W,\gamma^*,t\right)-\bar{p}^W\right) f(t)\,dt\right) \frac{dK^C}{d\bar{p}^W} + \int_{\hat{t}_0(K^C(\bar{p}^W),\bar{p}^W)}^{+\infty} \left(\frac{\partial \mathcal{S}}{\partial p} - v\left(\bar{p}^W,\gamma^*,t\right)\frac{\partial \mathcal{D}}{\partial p}\right) f(t)\,dt. \quad (6)$$

Again, the term multiplying $\frac{dK^C}{d\bar{p}^W}$ and the second integral are positive.

The analysis above yields the following:

Proposition 1 *Increasing the cap from the capacity-maximizing cap always increases welfare. If $K^C(.)$ is globally concave, the unique capacity-maximizing cap is lower than the welfare-maximizing cap, which is strictly lower than p^∞.*

Proof. *As seen previously, the capacity-maximizing cap* $\hat{p}^W$ *verifies* $\frac{dK^C}{d\bar{p}^W}\left(\hat{p}^W\right) = 0$*. Then,*

$$\frac{dW}{d\bar{p}^W}\left(\hat{p}^W\right) = \int_{\hat{t}_0(K^C,\hat{p}^W)}^{+\infty} \left(\frac{\partial \mathcal{S}}{\partial p} - v\left(\hat{p}^W, \gamma^*, t\right) \frac{\partial \mathcal{D}}{\partial p}\right) f(t)\, dt > 0.$$

Thus, a marginal increase in the cap always increases welfare.

Since $W\left(K^C(p), p\right)$ *is continuous on the compact set* $[(c+r), p^\infty]$*, a welfare maximizing cap* $\bar{p}^{W*}$ *exists. If* $K^C(.)$ *is globally concave,* $K^C(.)$ *hence* $W(.)$ *are increasing on the left of* $\hat{p}^W$*.* $\bar{p}^{W*}$ *cannot be lower than* $\hat{p}^W$*. Finally,*

$$\frac{dW}{d\bar{p}^W}(p^\infty) = \frac{dW_1}{d\bar{p}^W}(p^\infty) = \left(\int_{\hat{t}(K^C(p^\infty),c)}^{+\infty} -\frac{K^C(p^\infty)}{N} P_q\left(K^C(p^\infty), t\right) f(t)\, dt\right) \frac{dK^C}{d\bar{p}^W} < 0,$$

thus p^∞ *cannot be welfare maximizing, therefore* $\bar{p}^{W*} < p^\infty$*.* ■

Proposition 1 extends the previous analyses of price caps, which have focussed on their impact on supply, and have ignored their impact on demand, in particular customers rationing. In other words, previous analyses have concentrated on the first term in $\frac{dW}{d\bar{p}^W}$ and ignored the second.

This second term matters for policy makers. In addition to modify investment incentives, imposing a price cap when demand varies and the good is non storable means to some customers must be curtailed. Curtailment reduces welfare, as consumers are willing to pay more when curtailed than when they are not.

A direct and counterintuitive consequence of Proposition 1 is that $K^C\left(\bar{p}^{W*}\right) < K^C\left(\hat{p}^W\right)$: to maximize welfare, lower capacity than the maximum must be installed. The intuition is that the gains from reducing curtailments exceed the gains from additional capacity.

Policy makers should therefore attempt to quantity the impact of rationing when deciding on the imposition of price caps.

Proposition 1 shows that imposing the optimal price cap increases welfare compared to the never binding cap, which justifies the use of price cap. As will see in Sections 4 and later, it does not always hold if a fraction of customers face constant prices.

Finally, it is worth observing that even the welfare-maximizing cap yields lower capacity than first-best K^*, uniquely defined by

$$\Psi_0(K^*, c) = \int_{\hat{t}_0(K^*,c)}^{+\infty} (P(K^*, t) - c) f(t) \, dt - r = 0. \tag{7}$$

The marginal social value of capacity is simply the on-peak expected difference between the price and the marginal operating cost. At the optimum, this marginal value equals the marginal investment cost.

Suppose first the price cap is binding on-peak. Inserting equation (7) into equation (3) yields

$$\begin{aligned}
\Psi_1\left(K^*, \bar{p}^W\right) &= \int_{\hat{t}(K^*,c)}^{\hat{t}_0\left(K^*,\bar{p}^W\right)} \left(P(K^*; t) + \frac{K^*}{N} P_q(K; t) - c\right) f(t) \, dt + \int_{\hat{t}_0\left(K^*,\bar{p}^W\right)}^{+\infty} \left(\bar{p}^W - c\right) f(t) \, dt \\
&\quad - \int_{\hat{t}_0(K^*,c)}^{+\infty} (P(K^*, t) - c) f(t) \, dt \\
&= -\int_{\hat{t}_0(K^*,c)}^{\hat{t}(K^*,c)} (P(K^*, c) - c) f(t) \, dt + \int_{\hat{t}(K^*,c)}^{\hat{t}_0\left(K^*,\bar{p}^W\right)} \frac{K^*}{N} P_q(K; t) f(t) \, dt \\
&\quad - \int_{\hat{t}_0\left(K^*,\bar{p}^W\right)}^{+\infty} \left(P(K^*, t) - \bar{p}^W\right) f(t) \, dt \\
&< 0.
\end{aligned}$$

Since $\Psi_1(.,.)$ is decreasing in its first argument, $K^C\left(\bar{p}^W\right) < K^*$ for all $\bar{p}^W$ binding on-peak. The same argument applies to caps binding off-peak. Thus $K^C\left(\bar{p}^{W*}\right) < K^*$.

Even the optimal cap yields lower investment than the first best, for two reasons: first, imperfect competition leads to strategic under-investment. This produces the first two terms in the equation above (that disappear is the cap is binding off-peak). However, even if competition is perfect, i.e., $N \to +\infty$ and $\hat{t}(.,.) \to \hat{t}_0(.,.)$, a cap reduces investment compared to the first best: when the cap is binding, producers value the good at the cap $\bar{p}^W$, and not its true underlying value $P(K, t)$. Since the latter is higher than the former when the cap is binding by construction, this yields under-investment.

4 Introducing constant price customers

Suppose now only a fraction $\alpha > 0$ of customers face and react to real time wholesale price ("price reactive" customers), while the remaining fraction $(1 - \alpha)$ of customers face constant price p^R in all states of the world ("constant price" customers). For example price reactive customers purchase directly from the wholesale spot markets, while retailers serve constant price customers. The economics of retailing in this setting are discussed for example by Borenstein and Holland (2005) and Léautier (2014).

The presence of constant price customers is an essential characteristic of electric power markets, hence the analysis presented below uses electricity market vocabulary.

The analysis also apply to other markets where a fraction of customers pay the same price for the good irrespective of market conditions. Consider health care provision. In many countries, most customers pay the same price for a medical procedure, whether demand is high or low compared to supply. For example, the cost of a receiving an injection of a influenza vaccine is the same in the fall, before the flu season, and in the spring, after the flu season. Rationing is organized through waiting times (Schut and Van de Ven (2005)).

Having consumers pay a constant price, independent of the state of the world, may be a technical imperative, or a policy decision. In the electric power industry, the impossibility to measure consumption in real time and inform users about real time prices, has lead to a large fraction of customers facing a constant price. This constraint is progressively lifted by the advent of smart meters. Yet, few countries to date have embraced real-time pricing, for fear of transferring price risk onto households. In the health care industry, other reasons, for example transaction costs, or an imperative of fairness, may justify the use of fixed prices.

This article is not discussing the appropriateness of fixed price policies. It examines how introducing constant price customers interacts with the imposition of price caps.

4.1 Constant price customers, selective curtailment, and residual inverse demand

To introduce the notation, suppose first no price cap imposed.

Since a fraction of customers does not react to real time price, there may be instances when the System Operator (SO) has no alternative but to curtail demand, i.e., to interrupt supply.

Assumption 3 *The SO has the technical ability to curtail "constant price" consumers while not curtailing "price reactive" customers.*

Assumption 3 holds only partially today in electricity markets: most SOs can only organize curtailment by geographical zones, and cannot differentiate by type of customer. It will hold fully in a few years, when "smart meters" are rolled out, as is mandated in most European countries and many US states.

Define $\rho(Q;t)$ the residual inverse demand curve with possible curtailment of constant price customers

$$\rho(Q;t) = P\left(\frac{Q-(1-\alpha)\mathcal{D}\left(p^R,\gamma^*;t\right)}{\alpha};t\right) \tag{8}$$

where γ^* is the optimal serving ratio in state t for production Q.

Price reactive customers face the wholesale spot price $\rho(Q,t)$, hence are never curtailed at the optimum. Off-peak, demand is low, and firms play a symmetric Cournot equilibrium with residual demand $\rho(Q,t)$. On-peak, demand is equal to installed capacity K, and the wholesale price is $\rho(K;t)$.

As long as $\rho(K;t) \leq v\left(p^R,1;t\right)$, constant price customers are not curtailed in state t. If $\rho(K;t) > v\left(p^R,1;t\right)$, then, $\gamma^* < 1$ is set to equalize constant price customers' $VoLL$ and the wholesale price:

$$v\left(p^R,\gamma^*;t\right) = \rho(K;t).$$

For production Q, define[2] $\bar{t}(Q)$ the first state of the world for which the price equal the $VoLL$:

$$\rho(Q;\bar{t}) = v\left(p^R,1;\bar{t}\right).$$

Under perfect competition, price equals marginal cost when $Q \leq K$, hence the $VoLL$ may be reached only on-peak. If demand is elastic enough, price may never reach the $VoLL$, in which case $\bar{t}(Q) \to +\infty$.

If competition is imperfect and residual demand is very inelastic (e.g., α very low), the price may reach the $VoLL$ off-peak. In that case, there exists $\bar{t}$ such that

$$\rho\left(Q^C(c,\bar{t}),\bar{t}\right) = v\left(p^R,1,\bar{t}\right).$$

For $t \geq \bar{t}$, as long as the market is off-peak, the SO sets the wholesale price at $v\left(p^R,1,t\right)$. No rationing

[2] $\bar{t}$ is a function of all the parameters of the residual demand function, in particular α and p^R. The notation $\bar{t}(Q)$ is used since the dependency on production Q is the most important. This simplification is used for all functions used in the analysis.

occurs, since there is spare capacity. On-peak is reached when

$$\alpha D\left(v\left(p^{R},1,t\right),t\right)+\left(1-\alpha\right)D\left(p^{R},1,t\right)=K\Leftrightarrow\rho\left(K,t\right)=v\left(p^{R},1,t\right)\Leftrightarrow t=\bar{t}\left(K\right).$$

On-peak, constant price customers are curtailed, and the price is the $VoLL$ for the optimal curtailment: $\rho\left(K;t\right)=v\left(p^{R},\gamma^{*};t\right)$.

I assume that the fixed price p^R is sufficiently high for the $VoLL$ to always be higher than the long-term marginal cost of production, i.e.,

$$v\left(p^{R},1,0\right)>\left(c+r\right),$$

otherwise constant price customers would never be served.

Additional assumptions on the inverse demand and rationing technology are required to ensure that $\rho\left(Q;t\right)$ is well-behaved, in particular satisfies Assumption 1. I use the following notation: if no rationing occurs, $\rho_q=\frac{1}{\alpha}P_q\left(\frac{Q-(1-\alpha)D\left(p^R;t\right)}{\alpha};t\right)$; when rationing occurs, $\rho_q=\frac{\partial v}{\partial K}=\frac{\partial v}{\partial\gamma}\frac{\partial\gamma^*}{\partial K}$.

Assumption 4 Properties of the inverse demand and rationing technology. *For all* $t\geq 0$*,* $Q\geq q\geq 0$*,* $\alpha\in(0,1]\,0$*,* $p^R>0$*, and* $\gamma\in(0,1]$*: (i) the marginal revenue does not increase as production increases,* $\rho_q\left(Q;t\right)+q\rho_{qq}\left(Q;t\right)\leq 0$*, (ii) the marginal revenue increases as the state of the world increases,* $\rho_t\left(Q;t\right)+q\left|\rho_{qt}\left(Q;t\right)\right|>0$*, (iii) the* $VoLL$ *does not increase as the serving ratio increases,* $\frac{\partial v}{\partial\gamma}\leq 0$ *and increases as the state of the world increases* $\frac{\partial v}{\partial t}>0$*, and (iv)* $\left(\alpha\left(\frac{\partial D}{\partial p}\frac{\partial v}{\partial t}+\frac{\partial D}{\partial t}\right)+(1-\alpha)\frac{\partial\mathcal{D}}{\partial t}\right)>0$.

If no rationing occurs, $P_t\left(Q,t\right)>0$ and $P_q\left(Q,t\right)<0$ are sufficient to guarantee that $\rho_t\left(Q;t\right)>0$, $\rho_q\left(Q;t\right)<0$.

Suppose now rationing occurs for $t\geq\bar{t}\left(K\right)$. As shown in Appendix C, $\frac{\partial v}{\partial\gamma}\leq 0$ guarantees that $\frac{\partial\gamma^*}{\partial K}>0$ and $\frac{\partial v}{\partial K}\leq 0$. Conditions (i) and (ii) ensure that the second derivatives of $v\left(p^R,\gamma^*;t\right)$ have the desired properties. Conditions (iii) and (iv) guarantee that the optimal serving ratio decreases as the state of the world increases: $\frac{\partial\gamma^*}{\partial t}<0$. If curtailment occurs in state $\bar{t}$, it also occurs in all states $t\geq\bar{t}$. Furthermore, price increases as the state of the world increases: $\frac{dv}{dt}>0$.

Assumption 4 holds for example if inverse demand is linear with constant slope: $P\left(q;t\right)=a\left(t\right)-bq$ where $b>0$ and $a^{'}\left(t\right)>0$, and rationing perfectly anticipated: $\mathcal{S}\left(p,\gamma;t\right)=\gamma S\left(p;t\right)$. If no rationing occurs,

$$\rho(Q,t) = \frac{a(t) - bQ - (1-\alpha)p^R}{\alpha} \tag{9}$$

is linear, hence satisfies conditions (i) and (ii).

Since rationing is anticipated,

$$v\left(p^R,\gamma,t\right) = \frac{S\left(p^R,t\right)}{D\left(p^R,;t\right)} = a(t) - b\frac{D\left(p^R,t\right)}{2} = \frac{a(t)+p^R}{2}. \tag{10}$$

$\frac{\partial v}{\partial \gamma} = 0$, and $\frac{\partial v}{\partial t} = \frac{a^{'}(t)}{2} > 0$, and $v\left(p^R,\gamma^*;t\right)$ satisfies conditions (i) to (v).

Under Assumption 4, previous derivations apply if we replace $P(Q,t)$ by $\rho(Q,t)$. Thus, with a slight abuse, I use the same notation. For example, the Cournot equilibrium output $Q^C(c,t)$ is defined by

$$\rho\left(Q^C,t\right) + \frac{Q^C}{N}\rho_q\left(Q^C,t\right) = c,$$

and $\hat{t}(K,c)$ by

$$Q^C\left(c,\hat{t}\right) = K \Leftrightarrow \rho\left(Q^C\left(c,\hat{t}\right),\hat{t}\right) + \frac{Q^C\left(c,\hat{t}\right)}{N}\rho_q\left(Q^C\left(c,\hat{t}\right),\hat{t}\right) = c.$$

4.2 Admissible price caps

Suppose now a cap $\bar{p}^W$ is imposed. The introduction of constant price customers also modifies the set of admissible caps. The minimum price cap is unchanged: $\bar{p}^W \geq (c+r)$.

The highest binding cap p^∞, which was the upper bound for admissible caps when $\alpha = 1$, is no longer suitable when $\alpha < 1$. Instead, the $VoLL$ v is the natural upper bound for $\bar{p}^W$. To see why, suppose $\bar{p}^W > v$. When the cap is binding on-peak, the wholesale price is set at $\bar{p}^W > v$. The *SO* must then curtail all constant price customers, since their $VoLL$ is lower than the opportunity cost of power. This is irrealistic.

To be admissible, a price cap must be lower than the $VoLL$. When the cap is binding, the *SO* curtails constant price customers, who pay fixed price p^R, before price reactive customers, who pay $\bar{p}^W$ as long as $v\left(\bar{p}^W,1,t\right) > v\left(p^R,\gamma^*,t\right)$ where γ^* is the optimal serving ratio for constant price consumers. This condition is always verified if rationing is perfectly anticipated, and I assume it holds at the equilibrium, i.e., that the equilibrium capacity and the elasticity of residual demand are high enough that price reactive customers need not be curtailed.

At this point, it is helpful to discuss how $VoLL$ is used in power markets. In this work, I assume the SO knows exactly the $VoLL$ in every state of the world. While this assumption is highly unrealistic in practice, it constitutes a useful analytical benchmark.

In reality, regulators, SOs and economists have little idea of the $VoLL$. Estimation is extremely difficult, because the $VoLL$ varies dramatically across customer classes, states of the world, and duration and conditions of outages. Estimates vary in an extremely wide range from 2 000 £$/MWh$ in the British Pool in the 1990s to 200 000 \$$/MWh$ (see for example, Cramton (2000)). In practice, the SO uses her best estimate of the average $VoLL$, and prioritizes curtailment by geographic zones (using criteria such as economic activity, political weight, network conditions, etc.), thus implementing a third best.

Both approaches produce downward sloping demand curves, and are analytically equivalent.

How do we formalize the condition $\bar{p}^W \leq v$? If the $VoLL$ is a number (e.g., $20,000$ €$/MWh$), all caps lower than this number are admissible.

In this article, the SO uses the true $VoLL$, derived from the underlying demand function. Suppose first the $VoLL$ is reached on-peak. A cap is admissible if and only if:

$$p \leq v\left(p^R, 1, \bar{t}\left(K^C(p)\right)\right). \tag{11}$$

Suppose now the $VoLL$ is reached off-peak. As previously seen, on-peak is reached for $t = \bar{t}(K)$. Therefore, cap binding on-peak is admissible if and only if it satisfies condition (11).

The natural candidate for maximum admissible cap is the smallest fixed point of $v\left(p^R, 1, \bar{t}\left(K^C(.)\right)\right)$, defined by

$$\Phi = v\left(p^R, 1, \bar{t}\left(K^C(\Phi)\right)\right). \tag{12}$$

I prove in Appendix D that Φ exists, and satisfies

$$\hat{t}_0\left(K^C(\Phi), \Phi\right) = \bar{t}\left(K^C(\Phi)\right).$$

To understand this result, observe that "the price cap is lower than the $VoLL$" is equivalent to "the price cap is reached before the $VoLL$", i.e., $\hat{t}_0\left(K^C(p), p\right) \leq \bar{t}\left(K^C(p)\right)$. For the maximum admissible cap, the price cap and the $VoLL$ are reached simultaneously.

4.3 Summary of constant price customers

It is helpful to summarize the changes created by the introduction of constant price customers. First, inverse demand $P(Q, t)$ is replaced by residual inverse demand $\rho(Q; t)$ in all expressions.

Second, when the wholesale price reaches the cap $\bar{p}^W$, only constant price customers are curtailed.

Third, the maximum admissible cap is Φ, is lower than or equal to p^∞.

Given that the cap is admissible, two situations may occur. First, the cap is reached off-peak. As long as $\alpha D\left(\bar{p}^W, t\right) + (1-\alpha) D\left(p^R, t\right) < K$, no curtailment occurs. On-peak, generation produces at capacity and curtailment of constant price consumers occurs.

Alternatively, the price cap may be reached on-peak. Then, for $t \leq \hat{t}_0\left(K, \bar{p}^W\right)$, generation produces at capacity and no curtailment occurs. For $t > \hat{t}_0\left(K, \bar{p}^W\right)$, curtailment of constant price consumers occurs.

Since the price cap is by construction lower than the $VoLL$, curtailment of constant price customers is driven by the imposition of the cap, and the serving ratio is determined by

$$\alpha D\left(\bar{p}^W, t\right) + (1-\alpha) \mathcal{D}\left(p^R, \mathring{\gamma}, t\right) = K.$$

Since price reactive customers do not face the opportunity cost of the good, the serving ratio is not "optimal".

5 Capacity and welfare maximizing price caps with constant price customers

The analysis presented in this Section extends previous results, in particular Zöttl (2011), in two directions. First it provides slightly more general sufficient conditions for the existence of a price maximizing cap. Second, and more importantly, it shows that, contrary to the analysis presented by Zöttl (2011), the price maximizing cap may lead to a lower welfare than no price cap. To simplify the notation, I use p and no longer $\bar{p}^W$ to represent the price cap.

5.1 Capacity maximizing cap

Analysis similar to the case $\alpha = 1$ yields:

Proposition 2 *Suppose*

$$\rho_t \left(\alpha D \left(c + r, 0\right) + \left(1 - \alpha\right) D \left(p^R, 0\right), 0\right) > r f \left(0\right), \tag{13}$$

and, for all $p \in [(c + r), \Phi]$,

$$2 + \frac{1}{h\left(\hat{t}_0\right)} \left(\frac{f'\left(\hat{t}_0\right)}{f\left(\hat{t}_0\right)} - \frac{\rho_{t^2}}{\rho_t} \right) \left(K^C\left(p\right), p\right) > 0, \tag{14}$$

where $h\left(t\right) = \frac{f(t)}{1 - F(t)}$ *is the hazard rate.*

If

$$\rho_t \left(K^C\left(\Phi\right), \hat{t}_0\left(K^C\left(\Phi\right), \Phi\right)\right) < \left(\Phi - c\right) h \left(\left(\hat{t}_0\left(K^C\left(\Phi\right), \Phi\right)\right)\right), \tag{15}$$

there exists a unique capacity maximizing cap $\hat{p} < \Phi$.

If condition (15) *is not met, the capacity maximizing cap is* $\hat{p} = \Phi$*, the maximum admissible cap. In this case,* Φ *is also the welfare maximizing cap, and if* $\frac{\partial v}{\partial K} = 0$*, equilibrium capacity is lower than if no cap was imposed.*

Proof. *The formal proof is presented in Appendix E. Since* $K^C\left(.\right)$ *is continuously differentiable on* $[(c + r), \Phi]$*, a maximum exists. Condition* (13) *is equivalent to* $\frac{dK^C}{dp}\left(c + r\right) > 0$*, and condition* (15) *is equivalent to* $\frac{dK^C}{dp}\left(\Phi\right) < 0$*. Thus, if condition* (15) *holds, there is at least one* $\hat{p} \in ((c + r), \Phi)$ *such that* $\frac{dK^C}{dp}\left(\hat{p}\right) = 0$*. Finally, condition* (14) *guarantees that* $\frac{d^2 K^C}{dp^2}\left(\hat{p}\right) < 0$*, hence there exists a unique capacity-maximizing cap* $\hat{p} \in ((c + r), \Phi)$*. This proves the first part of the Proposition.*

If condition (15) *does not hold, then* $\frac{dK^C}{dp}\left(p\right) \geq 0$ *for all* $p \in [(c + r), \Phi]$*: the unique capacity-maximizing cap is* $\hat{p} = \Phi$.

From Proposition 1, $p^* = \Phi$ *since* $\frac{dW}{dp}\left(p\right) > 0$ *for all* $p \in [(c + r), \Phi]$*. It remains to prove that* $K^C\left(\Phi\right) \leq K^C_\infty$*. Suppose for example the* $VoLL$ *is reached off-peak, Appendix A proves* K^C_∞ *is defined by*

$$\Psi_4\left(K^C_\infty\right) = \int_{\bar{t}\left(K^C_\infty\right)}^{+\infty} \left(v\left(p^R, \gamma^*, t\right) + \frac{K}{N} v_q\left(p^R, \gamma^*, t\right) - c\right) f\left(t\right) dt - r = 0.$$

The equilibrium capacity for the maximum admissible cap Φ is defined by

$$\Psi_2\left(K^C(\Phi),\Phi\right)=\int_{\hat{t}_0(K^C(\Phi),\Phi)}^{+\infty}(\Phi-c)\,f(t)\,dt-r=0.$$

Then, if $v_q=0$,

$$\begin{aligned}\Psi_4\left(K^C(\Phi)\right) &= \int_{\bar{t}(K^C(\Phi))}^{+\infty}\left(v\left(p^R,\gamma^*,t\right)-c\right)f(t)\,dt-\int_{\hat{t}_0(K^C(\Phi),\Phi)}^{+\infty}(\Phi-c)\,f(t)\,dt\\ &= \int_{\bar{t}(K^C(\Phi))}^{+\infty}\left(v\left(p^R,\gamma^*,t\right)-\Phi\right)f(t)\,dt>0\end{aligned}$$

since $\hat{t}_0\left(K^C(\Phi),\Phi\right)=\bar{t}\left(K^C(\Phi)\right)$. Thus, $K^C(\Phi)<K^C_\infty$. A similar analysis proves the result holds if the $VoLL$ is reached on-peak. ■

Proposition 2 extends Zöttl (2011) derivation of the capacity-maximizing cap in multiple directions. First, it covers the realistic case of a fraction of constant price customers. Second, it derives the capacity-maximizing cap when the cap is reached off-peak. This is an important extension, as this case seems fairly realistic as long as a small fraction is price responsive. Third, it provides a more general sufficient condition for the existence of a unique capacity-maximizing cap, as Zöttl (2011) had limited his analysis to separable demands $P(Q,t)=t-\tilde{P}(Q)$.

Finally, Proposition 2 shows that, while a capacity-maximizing cap always exists if the sufficient conditions are met, this cap may lead to lower capacity than the absence of cap. This stands in sharp contrast with Zöttl (2011) Theorems 4 and 6. The intuition is that introduction of constant price customers imposes that the cap be lower than the $VoLL$, and not the highest binding cap. When condition (15) is not met, a marginal increase in the cap, if it was feasible, would lead to higher investment. This contrasts with the case $\alpha=1$, where a *reduction* in the cap from the highest binding cap leads to higher investment. Then, if $v_q=0$, the never binding cap leads to higher investment.

5.2 Welfare with and without a cap

We now compare the welfare with and without a cap.

Proposition 3 *If the welfare maximizing cap yields lower investment than no cap, imposing a price cap reduces welfare.*

Proof. *The details of the proof are presented in Appendix F. As shown in Section 3, introducing a cap has an impact on installed capacity and on rationing. Appendix F shows that, for $K \leq K_{\infty}^{C}$, an increase in capacity increases welfare. Second, the intuition of Proposition 1 extends to the presence of constant price customers: a cap increases rationing, hence reduces gross surplus. This then yields the result.* ■

Proposition 3 holds in particular if $p^* = \Phi$ and $v_q = 0$, but holds as long as $K^C(p^*) \leq K_{\infty}^{C}$. It reverses an important result. When a share of customers face a fixed price, no cap may lead to higher welfare than even the welfare-maximizing cap. This result has a "contradictory instruments" flavor. Imposing a fixed price for a fraction of customers and a cap on price reactive customers may be perceived as protecting customers against exercise of market power. Yet, in some instances, the interaction between the fixed price and the price cap leads to lower overall welfare than if only fixed prices were imposed.

6 Illustration on a specific case

The previous analysis indicates that, if condition (15) is not met, imposing a price cap reduces welfare. In this Section, I examine, for an illustrative example, the range of validity of condition (15).

6.1 Model specification

Suppose that (i) demand is linear with constant slope $P(Q,t) = a(t) - bQ$, and $a(t) = a_0 - a_1 e^{-\lambda_2 t}$, (ii) states of the world are distributed according to $f(t) = \lambda_1 e^{-\lambda_1 t}$, and (ii) rationing is perfectly anticipated. This specification provides an adequate representation of actual demand, while leading to closed-form expressions.

This specification satisfies condition (14). The hazard rate of the exponential distribution is $h(t) = \lambda_1 = -\frac{f'(t)}{f(t)}$. Then,

$$\rho_t = \frac{a_1 \lambda_2 e^{-\lambda_2 t}}{\alpha} \Rightarrow \rho_{t^2} = -\lambda_2 \rho_t \Leftrightarrow \frac{\rho_{t^2}}{\rho_t} = -\lambda_2.$$

Thus,

$$2+\frac{1}{h\left(\hat{t}_0\right)}\left(\frac{f'\left(\hat{t}_0\right)}{f\left(\hat{t}_0\right)}-\frac{\rho_{t^2}}{\rho_t}\right)=2+\frac{1}{\lambda_1}\left(-\lambda_1+\lambda_2\right)=1+\frac{\lambda_2}{\lambda_1}=\frac{\lambda+1}{\lambda}>0.$$

6.2 Critical thresholds

Conditions (1) and (11) are naturally expressed as functions of p. However, for the numerical implementation, we fix p and let K vary to determine $K^c(p)$. Therefore, it is helpful to recast these conditions as functions of K. The first on-peak state of the world $\hat{t}(K,c)$ is defined by

$$a\left(\hat{t}\right)=\alpha c+(1-\alpha)\,p^R+\frac{N+1}{N}bK,$$

the first on-peak state of the world for which the cap is binding $\hat{t}_0(K,p)$ by

$$a\left(\hat{t}_0\right)=\alpha p+(1-\alpha)\,p^R+bK,$$

and the first on-peak state of the world for which price reaches $VoLL$ $\bar{t}(K)$ by

$$a\left(\bar{t}\left(K\right)\right)=\frac{2}{2-\alpha}bK+p^R.$$

The price cap is reached on-peak (condition (1)) if and only if

$$\hat{t}\left(K,c\right)\leq\hat{t}_0\left(K,p\right)\Leftrightarrow bK\leq\alpha N\left(p-c\right)=bK_2\left(p,\alpha,N\right).$$

For a given p, the profit function and the marginal value of capacity are defined piecewise for $K\leq K_2$ and $K>K_2$.

The price cap is admissible, i.e., the price cap is lower than the $VoLL$ (condition (11)) if and only if

$$p\leq v\left(p^R,1,K^C\left(p\right)\right)\Longleftrightarrow\hat{t}_0\left(K^C\left(p\right),p\right)\leq\bar{t}\left(K^C\left(p\right)\right)$$

$\Leftrightarrow$

$$\alpha p+(1-\alpha)\,p^R+bK^C\left(p\right)\leq\frac{2}{2-\alpha}bK^C\left(p\right)+p^R\Leftrightarrow bK^C\left(p\right)\geq\left(2-\alpha\right)\left(p-p^R\right)=bK_3\left(\alpha,p\right).$$

Admissibility of the price cap provides a lower bound for the admissible values of $K^C(p)$.

For the values of c, p^R, α and N estimated (and discussed below), $K_3(\alpha, p) > K_2(p, \alpha, N) > K_4(\alpha, N)$ for all possible values of p. In particular, if the price cap is admissible, $K^C(p) \geq K_3(\alpha, p)$, then the cap is reached off-peak, $K^C(p) \geq K_2(p, \alpha, N)$. This simplifies the analysis, as shown below.

6.3 Equilibrium capacity

Since the price cap is binding off-peaking,

$$\Psi(K,p) = \Psi_2(K,p) = \int_{\hat{t}_0(K,p)}^{+\infty} (p-c) f(t)\, dt = (p-c)\left(e^{-\lambda_2 \hat{t}_0(K,p)}\right)^{\lambda} - r.$$

Equation (3) yields

$$(p-c)\left(\frac{a_0 - bK^C(p) - (1-\alpha)p^R - \alpha p}{a_1}\right)^{\lambda} = r$$

$\Leftrightarrow$

$$bK^C(p) = a_0 - p^R - \alpha\left(p - p^R\right) - a_1\left(\frac{r}{p-c}\right)^{\frac{1}{\lambda}}.$$

This immediately yields

$$b\frac{dK^C(p)}{dp} = -\alpha + \frac{1}{\lambda}\frac{a_1 r^{\frac{1}{\lambda}}}{(p-c)^{\frac{\lambda+1}{\lambda}}}$$

and

$$b\frac{d^2K^C\left(\bar{p}^W\right)}{dp^2} = -\frac{\lambda+1}{\lambda^2}\frac{a_1 r^{\frac{1}{\lambda}}}{(p-c)^{\frac{2\lambda+1}{\lambda}}} < 0.$$

$K^C(p)$ is globally concave. Since $\lim_{p\to+\infty}\frac{dK^C(p)}{dp} = -\alpha < 0$, if

$$b\frac{dK^C(c+r)}{dp} = \frac{a_1}{\lambda r} - \alpha < 0,$$

there exists a unique capacity-maximizing cap $\mathring{p} > (c+r)$ defined by

$$\mathring{p} = c + \left(\frac{a_1 r^{\frac{1}{\lambda}}}{\alpha\lambda}\right)^{\frac{\lambda}{1+\lambda}}.$$

Assuming $\mathring{p}$ is admissible, i.e., $\mathring{p} \leq \Phi$, $\mathring{p}$ is increasing in c and r, increasing in a_1, and decreasing α. A more inelastic demand, either due to lower underlying elasticity (higher a_1) or to lower share of price reactive customers (lower α), leads to a higher capacity maximizing cap.

6.4 Maximum admissible price cap Φ

After algebraic manipulations, the function $g(p)$ becomes

$$g(p) = p - v\left(p^R, 1, \bar{t}\left(K^C(p)\right)\right) = \frac{2p + a_1\left(\frac{r}{p-c}\right)^{\frac{1}{\lambda}} - a_0 - p^R}{2-\alpha}.$$

Therefore,

$$g'(p) = \frac{1}{2-\alpha}\left(2 - \frac{a_1}{\lambda}\left(\frac{r}{p-c}\right)^{\frac{1}{\lambda}}\frac{1}{p-c}\right).$$

We verify that $g(.)$ decreasing then increasing. Since $lim_{p\to+\infty} g(p) = +\infty$, if $g(c+r) = \frac{2(c+r)+a_1-a_0-p^R}{2-\alpha} < 0$, Φ is uniquely defined by

$$g(\Phi) = 0 \Longleftrightarrow \Phi + \frac{a_1}{2}\left(\frac{r}{\Phi - c}\right)^{\frac{1}{\lambda}} = \frac{a_0 + p^R}{2}. \tag{16}$$

Observe that

$$\frac{a_0 + p^R}{2} = \lim_{t\to+\infty} v\left(p^R, t\right) = \lim_{t\to+\infty} \rho\left(K^C_\infty, t\right) = p^\infty.$$

Equation (16) illustrates that $\Phi < p^\infty$. Equation (16) also indicates that Φ does not depend on α.

Since equation (16) does not provide a closed-form expression of Φ, I use the equivalent condition $g(\mathring{p}) \leq 0$ to obtain a closed-from expression of the condition $\mathring{p} \leq \Phi$:

$$g(\mathring{p}) \leq 0 \Leftrightarrow \mathring{p} + \frac{a_1}{2}\left(\frac{r}{\mathring{p} - c}\right)^{\frac{1}{\lambda}} \leq \frac{a_0 + p^R}{2}.$$

Observing that

$$\frac{a_1}{2}\left(\frac{r}{\mathring{p} - c}\right)^{\frac{1}{\lambda}} = \frac{1}{2} r^{\frac{1}{1+\lambda}} a_1^{\frac{\lambda}{1+\lambda}} (\alpha\lambda)^{\frac{1}{1+\lambda}},$$

algebraic manipulations yield

$$\mathring{p} \leq \Phi \Leftrightarrow c + \left(1 + \frac{\lambda\alpha}{2}\right)\left(\frac{a_1 r^{\frac{1}{\lambda}}}{\alpha\lambda}\right)^{\frac{\lambda}{1+\lambda}} \leq \frac{a_0 + p^R}{2}. \tag{17}$$

As α decreases, the left-hand side of condition (17) increases, which renders the condition more difficult to meet. This confirms previous intuition: lowering α increases $\mathring{p}$ while not affecting Φ.

The impact of underlying demand elasticity η on condition (17) is ambiguous, since both a_1 on the left hand-side and a_0 on the right-hand side are impacted by change in η.

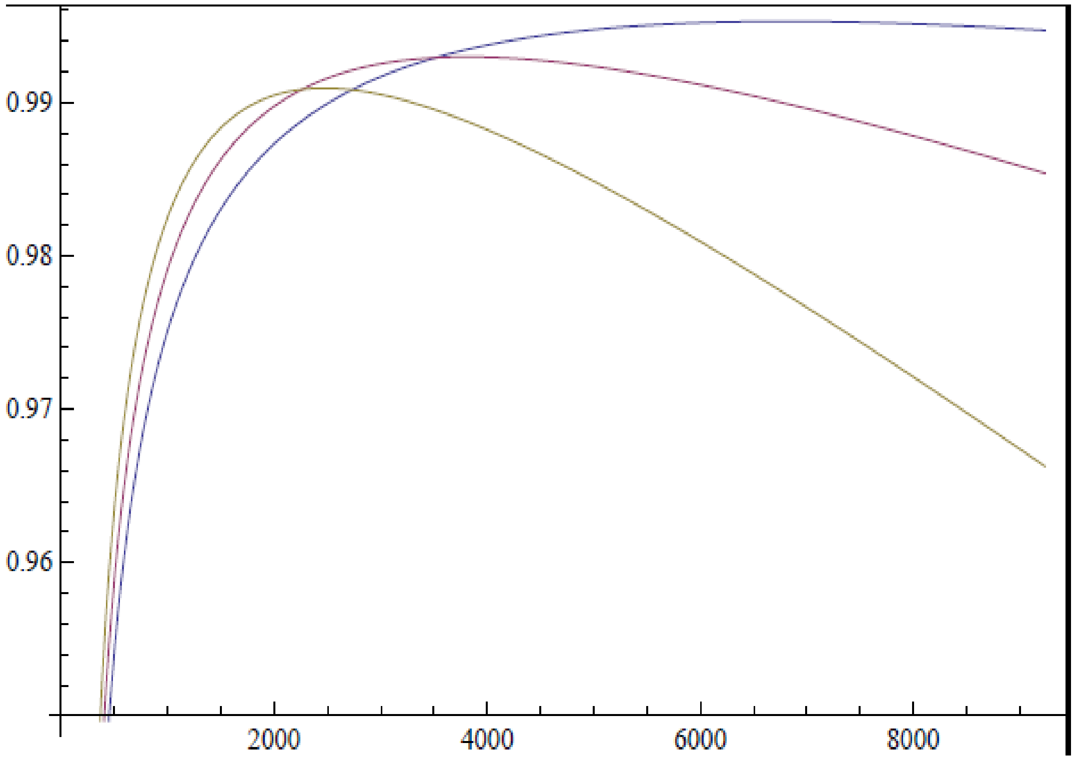
6.5 Numerical illustration

I estimate the parameters of the model using the realizations of electricity demand for every half-hour in France in 2010, a set of data known as a load duration curve. a_0, a_1, λ, and bQ^{∞}, where $Q^{\infty} = \frac{a_0 - p_0}{b}$ is the maximum demand for price p_0, are the parameters to be estimated. λ is estimated by Maximum Likelihood. The same load duration curve provides an expression of a_0 and a_1 as a function of bQ^{∞}. The average demand elasticity η is then used to estimate bQ^{∞}. Of course, estimates of the short-run elasticity of demand are very uncertain. I test two values of elasticities, $\eta = -0.01$ and $\eta = -0.1$ at price $p_0 = 100$ €$/MWh$, that correspond respectively to the lower bound and upper bound of estimates reported by Lijesen (2007). Following this procedure, Léautier (2014) estimates

$$\begin{matrix} for\ \eta = -0.1 & & for\ \eta = -0.01 \\ \begin{cases} bQ^{\infty} = 1\ 873\ €/MWh \\ a_0 = 1\ 973\ €/MWh \\ a_1 = 1\ 236\ €/MWh \\ \lambda = 1.78 \end{cases} & and & \begin{cases} bQ^{\infty} = 18\ 727\ €/MWh \\ a_0 = 18\ 827\ €/MWh \\ a_1 = 12\ 360\ €/MWh \\ \lambda = 1.78 \end{cases} \end{matrix}.$$

Investment and operating costs are those of a Gas Combustion Turbine, as provided by the International Energy Agency (median case, IEA (2010)): $c = 72$ €$/MWh$ and $r = 6$ €$/MWh$. The regulated price is $p^R = 50$ €$/MWh$, close to the average of the energy component electricity price in most European markets[3]. For simplicity, network charges, retail margins and taxes are excluded from the analysis, as they vary across customer classes. Industry experts suggest $\alpha = 2\%$ is a lower bound for the current share of price reactive customers in most market, and $\alpha = 10\%$ would be an upper bound. As will be shown below, the price cap is reached before firms produce at capacity, thus the number of firms does not matter.

[3]Eurostat, Table 2 Figure 2 from http://epp.eurostat.ec.europa.eu/statistics_explained/images/a/a1/Energy_prices_2011s2.xls



6.5.1 Low price elasticity

Consider first the case $\eta = -0.01$. Equation (16) is solved numerically to yield $\Phi = 9,237$ €$/MWh$.

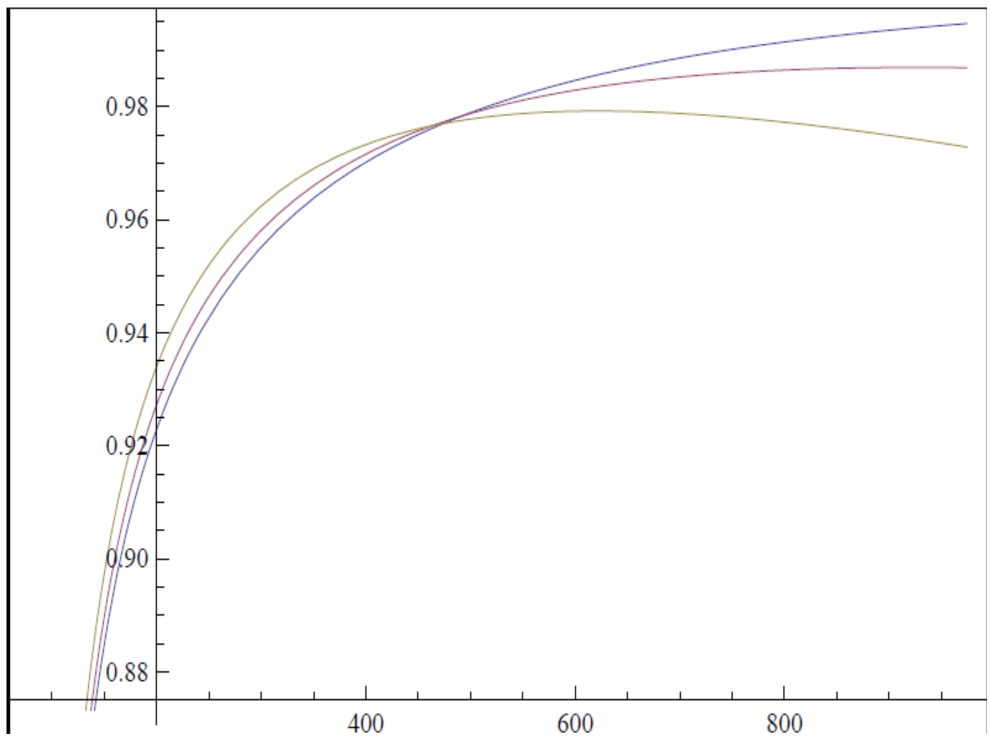
Since α vary, I express the Cournot capacity as a percentage of the first-best capacity. $K^C(p)/K^*(\alpha)$ for $p \in [c+r, \Phi]$ is presented below for $\alpha = 2\%$ (blue line on top), $\alpha = 5\%$ (purple line intermediate), and $\alpha = 10\%$ (brown line at the bottom).

The capacity-maximizing cap p^* and the maximum capacity $K^C(p^*)/K^*(\alpha)$ for each value of α, are:

α (%)	2	5	10
p^* (€$/MWh$)	$6,793$	$3,810$	$2,470$
$K^C(p^*)/K^*(\alpha)$ (%)	98.3	97.5	96.7

For all values of α, there exists a unique interior capacity-maximizing cap, i.e., $p^* < \Phi$. Furthermore, Cournot capacity at p^* is very close to the first-best.

The optimal price cap p^* is higher than the caps in place in most US markets, and except for the case $\alpha = 10\%$, higher than the $3,000$ €$/MWh$ cap on the day-ahead market in continental Europe and the $5,000$ $\$/MWh$ cap in effect in Texas.



6.5.2 High price elasticity

Consider now $\eta = -0.1$. Equation (16) is solved numerically to yield $\Phi = 974$ €$/MWh$. Why is $\Phi = 974$ €$/MWh$ much lower than the caps in effect in Europe and in Texas? Policy makers use their estimate of the $VoLL$ to set the cap, and not the $VoLL$ consistent with the underlying demand function. Thus, their estimate of the $VoLL$ is consistent with demand elasticity of $\eta = -0.01$ (at price 100 €/MWh, assuming demand is linear), and inconsistent with demand elasticity of $\eta = -0.1$. This low elasticity bias is consistent with the view held by many policy makers and commentators that electricity is an essential good, and that disruptions are very costly.

$K^C\left(\bar{p}^W\right)/K^*(\alpha)$ is presented below $\alpha = 2\%$ (blue line on top), $\alpha = 5\%$ (purple line intermediate), and $\alpha = 10\%$ (brown line at the bottom).

The capacity maximizing price cap p^* and the maximum capacity are:

α (%)	2	5	10
p^* (€$/MWh$)	974	927	621
$K^C(p^*)/K^*(\alpha)$ (%)	97.7	96.3	94.4

.

For $\alpha = 2\%$, $p^* = \Phi$: the capacity cap is the highest admissible cap. Thus, Cournot capacity is strictly lower than without a cap, although the numerical difference is small. For higher values of α, $p^* < \Phi$.

The resulting Cournot capacity is slightly further away from the first-best.

7 Concluding remarks

Price caps are often used by policy makers to "regulate markets". Previous analyses have focussed on the "supply side" impact of these caps, and derived the optimal price cap, which maximizes investment and welfare. This article expands the analysis to include the "demand side" impact of price caps: when prices can no longer rise, customers must be rationed to adjust demand to available supply. This yields two new findings, that contradict previous analyses. First, the welfare-maximizing cap is higher than the capacity-maximizing cap, since increasing the cap increases gross surplus when customers are rationed. Second, in some cases, the capacity-maximizing cap leads to lower capacity and welfare than no cap. These findings underscores the importance for policy makers to examine the impact on customers when they impose price caps.

These findings are particularly relevant for the electric power industry where caps are frequently imposed to limit the exercise of market power. They also apply to other industries, such as housing, where rent controls are often imposed, and health care, where a fraction of customers face constant prices, and caps on prices are sometimes imposed.

It would be important to test empirically these findings. An illustrative model of the power industry shows that no price cap producing higher welfare than any cap arises for realistic values of the parameters. I would like to test this hypothesis using a different specification and demand for the power industry. Closed-form solutions may not obtain, but numerical analysis will confirm (or infirm) the findings. It would also be interesting to test other industries, for example housing and health care, where price controls are prevalent, and market power, at least in some local markets, a possibility.

References

[1] M. Boiteux. La tarification des demandes en pointe: Application de la théorie de la vente au cout marginal. *Revue Générale d'Electricité*, 58:321–40, 1949.

[2] S. Borenstein and S. Holland. On the efficiency of competitive electricity markets with time-invariant retail prices. *RAND Journal of Economics*, 36:469–493, 2005.

[3] M. Crew and P. Kleindorfer. Peak load pricing with a diverse technology. *The Bell Journal of Economics*, 7(1):207–231, Spring 1976.

[4] L. Evans and G. Gunthrie. Price-cap regulation and the scale and timing of investment. *RAND Journal of Economics*, 43(3):537–561, Fall 2012.

[5] IEA. *Projected Cost of Generating Electricity.* OECD/IEA, 2010.

[6] P. Joskow and J. Tirole. Reliability and competitive electricity markets. *RAND Journal of Economics*, 38(1):60–84, Spring 2007.

[7] P. L. Joskow. Competitive electricity markets and investment in new generating capacity. in: Helm, D. (Ed.), The New Energy Paradigm, Oxford University Press, 2007.

[8] S. Késenne. The impact of salary caps in professional team sports. *Scottish Journal of Political Economy*, 47(4):422–430, September 2000.

[9] T.-O. Léautier. Is mandating smart meters smart? *The Energy Journal*, Forthcoming 2014.

[10] M. Lijesen. The real-time price elasticity of electricity. *Energy Economics*, 29:249–258, 2007.

[11] M. Mougeot and F. Naegelen. Hospital price regulation and expenditure cap policy. *Journal of Health Economics*, 24:55–72, 2005.

[12] F. Roques and N. Savva. Investment under uncertainty with price ceilings in oligopolies. *Journal of Economic Dynamics and Control*, 33(2):507–524, February 2009.

[13] F. T. Schut and W. P. V. de Ven. Rationing and competition in the dutch health care system. *Health Economics*, 14(S1):S59–S74, September 2005.

[14] S. Stoft. *Power System Economics.* Wiley-Interscience, 2002.

[15] G. Zöttl. On optimal scarcity prices. *International Journal of Industrial Organization*, 29(5):589–605, September 2011.

A Proof of Lemma 1

Consider N producers, each with installed capacity k^n, ordered such that $k^1 \leq ... \leq k^N$. $\Pi^n \left(k^n, \mathbf{k}^{-n}\right)$ is producer's n profit for the two-stage game.

A.1 Price cap binding on-peak

Suppose the price cap $\bar{p}^W$ is binding on-peak: $\hat{t}(K, c) \leq \hat{t}_0 \left(K, \bar{p}^W\right)$. Technical supplement G derives the equilibrium capacity in that case. Lemma 2 presented in Technical supplement G characterizes the equilibrium of the energy markets as follows: there exists N critical states of the world $0 \leq t^1 \leq t^2 \leq ... \leq t^N$ such that, for $t \in [t^j, t^{j+1}]$, all producers $i \leq j$ produce their entire capacity k^i, while all producers $n > j$ produce $\phi^{j+1}(k^1, ..., k^j; t)$ defined on $[t^j, t^{j+1}]$ as the solution of a "modified Cournot condition":

$$\rho \left(\sum_{i=1}^{j} k^i + (N-j)\, \phi^{j+1}; t \right) - c + \phi^{j+1} \rho_q \left(\sum_{i=1}^{j} k^i + (N-j)\, \phi^{j+1}; t \right) = 0.$$

$\widehat{Q}(t) = \sum_{i=1}^{j} k^i + (N-j)\, \phi^{j+1}(t)$ is increasing in t. For $t \leq t^1$, $\widehat{Q}(t) = Q^C(c, t)$. For $t \geq t^N$, all producers produce their entire capacity k^n, hence $\widehat{Q}(t) = \sum_{n=1}^{N} k^n$.

A.2 Price cap binding off-peak

Suppose the price cap is reached off-peak: $\hat{t}_0 \left(K, \bar{p}^W\right) < \hat{t}(K, c)$. Without loss of generality since we are ultimately looking for a symmetric equilibrium, suppose the price cap is binding before t^1, i.e., there exists $\tilde{t}^0 \in \left[0, t^1\right]$ such that $\rho \left(\widehat{Q}\left(\tilde{t}^0\right); \tilde{t}^0\right) = \bar{p}^W$.

Define $\tilde{q}^1(t)$ by $\rho \left(N \tilde{q}^1(t); t\right) = \bar{p}^W$. $q^n = \tilde{q}^1(t)$ for all $n \geq 1$ is the unique equilibrium for $t \geq \tilde{t}^0$, as long as producer 1 is not constrained. To prove the result, it is helpful to ensure that the Cournot price $\rho \left(\widehat{Q}(t); t\right)$ is increasing with the state of the world. This requires a slightly stronger Assumption than Assumption 4:

Assumption 5 *For all* $t \geq 0$, $Q \geq 0$, $\alpha \in (0, 1]\, 0$, *and* $p^R > 0$,

$$\rho_t(Q, t) \left(1 + \frac{Q \rho_{qq}(Q, t)}{\rho_q(Q, t)}\right) > Q \rho_{qt}(Q, t).$$

Then, $\forall t > \tilde{t}^0$

$$\rho\left(\widehat{Q}(t);t\right) > \rho\left(\widehat{Q}\left(\tilde{t}^0\right);\tilde{t}^0\right) = \bar{p}^W,$$

hence

$$\rho\left(N\tilde{q}^1(t);t\right) = \bar{p}^W < \rho\left(\widehat{Q}(t);t\right)$$

$\Longleftrightarrow$

$$\tilde{q}^1(t) > \frac{\widehat{Q}(t)}{N}.$$

This is the classical result: in each state of the world, a price cap reduces firms' ability to reduce output.

Suppose all firms $n > 1$ produce $\tilde{q}^1(t)$, while firm 1 considers deviating. A negative deviation is unprofitable since it reduces output but cannot increase price, which is capped at $\bar{p}^W$. Then, since $\tilde{q}^1(t) > \frac{\widehat{Q}(t)}{N}$ and the marginal revenue is decreasing,

$$\begin{aligned}\frac{\partial\pi^1}{\partial q^1}\left(\tilde{q}^1(t),...,\tilde{q}^1(t);t\right) &= \rho\left(N\tilde{q}^1(t);t\right) + \tilde{q}^1(t)\rho_q\left(N\tilde{q}^1(t);t\right) - c \\ &< \rho\left(\widehat{Q}(t);t\right) + \frac{\widehat{Q}(t)}{N}\rho_q\left(\widehat{Q}(t);t\right) - c = \frac{\partial\pi^n}{\partial q^n}\left(\frac{\widehat{Q}(t)}{N},...,\frac{\widehat{Q}(t)}{N};t\right) = 0.\end{aligned}$$

A positive deviation is not profitable. $q^n = \tilde{q}^1(t)$ for all $n \geq 1$ is a symmetric equilibrium. Since the profit function is concave, this equilibrium is unique.

When $t = \tilde{t}^1$ characterized by $\tilde{q}^1(t) = k^1 \Longleftrightarrow \rho\left(Nk^1;t^1\right) = \bar{p}^W$, producer 1 is constrained.

Similarly, for $t \in \left[\tilde{t}^j, \tilde{t}^{j+1}\right]$ for $j = 1,...,N-1$, the unique symmetric equilibrium for the $(N-j)$ remaining producers is $\tilde{q}^{j+1}(t)$ where $\rho\left(\sum_{i=1}^{j} k^i + (N-j)\tilde{q}^{j+1}(t);t\right) = \bar{p}^W$. $\tilde{t}^N$ is such that $\rho\left(\sum_{j=1}^{N} k^j;\tilde{t}^N\right) = \bar{p}^W$, hence $\tilde{t}^N = \hat{t}_0\left(K,\bar{p}^W\right)$ previously defined. For $t > \tilde{t}^N$, since wholesale price is fixed at $\bar{p}^W$ and generation is at capacity, the *SO* must curtail constant price consumers.

Consider now producer n expected profit, given capacities $k^1 \leq ... \leq k^N$, and the structure of the equilibrium described above:

$$\begin{aligned}\Pi^n\left(k^n;\mathbf{k}_{-n}\right) &= \int_0^{\tilde{t}^0}\frac{\hat{Q}}{N}\left(\rho\left(\hat{Q}\right)-c\right)f\left(t\right)dt \\ &+\left(\bar{p}^W-c\right)\left(\sum_{i=0}^{n-1}\int_{\tilde{t}^i}^{\tilde{t}^{i+1}}\tilde{q}^{i+1}\left(t\right)f\left(t\right)dt+k^n\left(1-F\left(\tilde{t}^n\right)\right)\right)-rk^n\end{aligned}$$

Thus, for all $n\leq N$,

$$\frac{\partial\Pi^n}{\partial k^n}=\int_{\tilde{t}^n}^{+\infty}\left(\bar{p}^W-c\right)f\left(t\right)dt-r$$

since output is continuous, and

$$\frac{\partial^2\Pi^n}{\left(\partial k^n\right)^2}=-\left(\bar{p}^W-c\right)f\left(\tilde{t}^n\right)\frac{\partial\tilde{t}^n}{\partial k^n}<0.$$

If there exists K such that

$$\frac{\partial\Pi^n}{\partial k^n}\left(\frac{K}{N},...,\frac{K}{N}\right)=\int_{\hat{t}_0\left(K,\bar{p}^W\right)}^{+\infty}\left(\bar{p}^W-c\right)f\left(t\right)dt-r=0,$$

then $\frac{K}{N}$ for all $n=1,...,N$ is the unique symmetric equilibrium. For any $t\geq 0$, $\lim_{K\to+\infty}\rho\left(K;t\right)<c$, thus $\lim_{K\to+\infty}\bar{t}\left(K,\bar{p}^W\right)=+\infty$, hence $\lim_{K\to+\infty}\frac{\partial\Pi^n}{\partial k^n}\left(\frac{K}{N},...,\frac{K}{N}\right)=-r<0$.

If $\rho\left(0,t\right)\geq\bar{p}^W$ for all $t\geq 0$, which I assume holds, $\hat{t}_0\left(0,\bar{p}^W\right)=0$ and

$$\frac{\partial\Pi^n}{\partial k^n}\left(0,...,0\right)=\int_0^{+\infty}\left(\bar{p}^W-c\right)f\left(t\right)dt-r=\bar{p}^W-\left(c+r\right)>0.$$

A.3 No price cap imposed

If the $VoLL$ is reached on-peak, the equilibrium is identical to the case $\alpha=1$, recognizing that $\rho\left(K,t\right)=v\left(p^R,\gamma^*,t\right)$ when curtailment occurs.

Suppose the $VoLL$ is reached off-peak: $\bar{t}\left(K\right)<\hat{t}\left(K,c\right)$. As before, suppose the $VoLL$ is reached before producer 1 reaches capacity, i.e. there exists $\widehat{t}^0\in\left[0,t^1\right]$ such that $\rho\left(\widehat{Q}\left(\widehat{t}^0\right);\widehat{t}^0\right)=v\left(p^R,1,\widehat{t}^0\right)$. As long as the total generation is not at capacity, there is no rationing. Price is thus $v\left(p^R,1,t\right)$, independent of output. The same argument as above shows that, $q^n=\hat{q}^1\left(t\right)$ defined by $\rho\left(N\hat{q}^1\left(t\right);t\right)=v\left(p^R,1,t\right)$ for all $n\geq 1$ is the unique equilibrium for $t\geq\widehat{t}^0$, as long as producer 1 is not at capacity.

Similarly, for $t\in\left[\hat{t}^j,\hat{t}^{j+1}\right]$ for $j=1,...,N-1$, the unique symmetric equilibrium for the $\left(N-j\right)$

remaining producers is $\hat{q}^{j+1}(t)$ where $\rho\left(\sum_{i=1}^{j} k^i + (N-j)\,\hat{q}^{j+1}(t)\,;t\right) = v\left(p^R,1,t\right)$.

Consider now producer n expected profit, given capacities $k^1 \leq ... \leq k^N$, and the structure of the equilibrium described above:

$$\begin{aligned}
\Pi^n\left(k^n;\mathbf{k}_{-n}\right) &= \int_0^{\hat{t}^0} \frac{\hat{Q}}{N}\left(\rho\left(\hat{Q}\right)-c\right) f(t)\,dt \\
&\quad + \left(v\left(p^R,1,t\right)-c\right)\left(\sum_{i=1}^{n-1}\int_{\hat{t}^i}^{\hat{t}^{i+1}} \hat{q}^{i+1}(t)\, f(t)\,dt\right) \\
&\quad + k^n\left(\int_{\hat{t}^n}^{\bar{t}(K)}\left(v\left(p^R,1,t\right)-c\right) f(t)\,dt + \int_{\bar{t}(K)}^{+\infty}\left(v\left(p^R,\gamma^*,t\right)-c\right) f(t)\,dt - r\right).
\end{aligned}$$

Thus, for all $n \leq N$,

$$\frac{\partial \Pi^n}{\partial k^n} = \int_{\hat{t}^n}^{\bar{t}(K)}\left(v\left(p^R,1,t\right)-c\right) f(t)\,dt + \int_{\bar{t}(K)}^{+\infty}\left(v\left(p^R,\gamma^*,t\right)+k^n\frac{\partial v}{\partial K}\left(p^R,\gamma^*,t\right)-c\right) f(t)\,dt - r$$

since output is continuous. If a symmetric equilibrium $\left(\frac{K^C}{N},...,\frac{K^C}{N}\right)$ exists, it satisfies:

$$\frac{\partial \Pi^n}{\partial k^n}\left(\frac{K^C}{N},...,\frac{K^C}{N}\right) = \int_{\bar{t}(K)}^{+\infty}\left(v\left(p^R,\gamma^*,t\right)+\frac{K^C}{N}\frac{\partial v}{\partial K}\left(p^R,\gamma^*,t\right)-c\right) f(t)\,dt - r = 0$$

As before, $lim_{\frac{K^C}{N}\to+\infty}\frac{\partial \Pi^n}{\partial k^n}\left(\frac{K^C}{N},...,\frac{K^C}{N}\right) < 0$ and $\frac{\partial \Pi^n}{\partial k^n}(0,...,0) = \int_0^{+\infty}\left(v\left(p^R,\gamma^*,t\right)-c\right) f(t)\,dt - r > \mathbb{E}\left[\rho(0,t)\right]-(c+r) > 0$, thus existence of $\frac{K^C}{N}$ is guaranteed.

Proving that $\left(\frac{K^C}{N},...,\frac{K^C}{N}\right)$ is the unique symmetric equilibrium is a bit more complex, since we need to treat separately an upward and a downward deviation. It is proven following the argument presented in the Technical supplement.

B Existence and characterization of the maximum of $K^C(p)$

Existence $\Psi_1(.,.)$ and $\Psi_2(.,.)$ are continuously differentiable in both variables by inspection; for any K, $\Psi(K,.)$ is continuous for at $\bar{p}^W$ such that $\hat{t}(K,c) = \hat{t}_0\left(K,\bar{p}^W\right)$, hence $K^C(.)$ is continuous on the compact set $\left[(c+r),p^\infty\right]$, thus, admits a maximum.

Furthermore, $K^C(.)$ is continuously differentiable. By inspection of equation (4), the only potential problem could arise at the pivotal price cap p_0 defined $\hat{t}\left(K^C(p_0), c\right) = \hat{t}_0\left(K^C(p_0), p_0\right)$. However,

$$\lim_{\substack{p\to p_0 \\ p\leq p_0}} \frac{\partial \Psi_1}{\partial K} = \frac{K^C(p_0)}{N} P_q\left(K^C(p_0); \hat{t}_0\right) f\left(\hat{t}_0\right) \frac{\partial \hat{t}_0}{\partial K} = -\left(p_0 - c\right) f\left(\hat{t}_0\right) \frac{\partial \hat{t}_0}{\partial K} = \lim_{\substack{p\to p_0 \\ p\geq p_0}} \frac{\partial \Psi_2}{\partial K}$$

since $p_0 + \frac{K^C(p_0)}{N} P_q\left(K^C(p_0); \hat{t}_0\right) = c$ by construction. Thus, $K^C(.)$ is continuously differentiable on $[(c+r), p^\infty]$.

The long-run marginal cost $(c+r)$ is not a maximum Let us examine $\frac{dK^C}{dp}(c+r)$. We first prove that a cap set at $(c+r)$ is binding off-peak. By contradiction, suppose the cap is reached on-peak:

$$\Psi_1(K, c+r) = \int_{\hat{t}(K,c)}^{\hat{t}_0(K,c+r)} \left(\rho(K;t) + \frac{K}{N}\rho_q(K;t) - c\right) f(t)\, dt + (c+r-c)\left(1 - F\left(\hat{t}_0(K, c+r)\right)\right) - r$$

For $t \leq \hat{t}_0(K, c+r)$, $\rho(K;t) \leq (c+r)$, thus $\rho(K;t) + \frac{K}{N}\rho_q(K;t) - c < r$, hence

$$\Psi(K, c+r) < r\left(1 - F\left(\hat{t}(K,c)\right)\right) - r \leq rF\left(\hat{t}(K,c)\right) \leq 0.$$

Thus, there cannot exists K^C such that $\Psi\left(K^C, c+r\right) = r$. This constitutes a contradiction, a cap set at $(c+r)$ is binding off-peak. By continuity, so is a cap set around $(c+r)$.

Since the cap is binding off-peak,

$$\Psi(K, c+r) = \Psi_2(K, c+r) = -rF\left(\hat{t}_0(K, c+r)\right).$$

Thus,

$$\Psi\left(K^C(c+r), c+r\right) = 0 \Leftrightarrow F\left(\hat{t}_0\left(K^C(c+r), c+r\right)\right) = 0 \Leftrightarrow \hat{t}_0\left(K^C(c+r), c+r\right) = 0.$$

On-peak starts at $t = 0$. Thus,

$$P\left(K^C(c+r), 0\right) = (c+r) \Leftrightarrow K^C(c+r) = D\left((c+r), 0\right).$$

Inserting into equation (4), and observing that

$$\frac{\partial \Psi_2}{\partial K}(K,p) = -(p-c) f\left(\hat{t}_0(K,p)\right) \frac{\partial \hat{t}_0(K,p)}{\partial K}, \frac{\partial \hat{t}_0(K,p)}{\partial p} = \frac{1}{P_t\left(K,\hat{t}_0\right)}, \ and \ \frac{\partial \hat{t}_0(K,p)}{\partial K} = \frac{-P_q\left(K,\hat{t}_0\right)}{P_t\left(K,\hat{t}_0\right)}$$

yields

$$\frac{dK^C}{dp}(c+r) = \frac{1 - rf(0)\frac{1}{P_t(K^C(c+r),0)}}{rf(0)\left(\frac{-P_q(K^C(c+r),0)}{P_t(K^C(c+r),0)}\right)} = \frac{P_t\left(K^C(c+r),0\right) - rf(0)}{-P_q\left(K^C(c+r),0\right) rf(0)},$$

thus

$$\frac{dK^C}{dp}(c+r) > 0 \Leftrightarrow P_t\left(D\left((c+r),0\right),0\right) - rf(0) > 0$$

which I assume holds.

The first never-binding cap p^∞ is not a maximum The Cournot capacity for the highest binding cap p^∞ is $K_\infty^C = K^C(p^\infty)$. Since p^∞ is (almost) never binding, it is not binding off-peak, hence K_∞^C defined by

$$\int_{\hat{t}(K_\infty^C,c)}^{+\infty} \left(P\left(K_\infty^C;t\right) + \frac{K}{N} P_q\left(K_\infty^C;t\right) - c\right) f(t)\,dt = r.$$

Introduce also

$$MR^\infty = \lim_{t\to+\infty}\left(\left(P\left(K_\infty^C,t\right) + \frac{K}{N}P_q\left(K_\infty^C;t\right)\right)\right).$$

$MR^\infty < p^\infty$ since $P_q(.,.) < 0$. Then for $\bar{p}^W \in (MR^\infty, p^\infty)$ and high enough that it is not binding off-peak, the marginal value of capacity at K_∞^C is

$$\begin{aligned}
\Psi_1\left(K_\infty^C,\bar{p}^W\right) &= \int_{\hat{t}(K_\infty^C,c)}^{\hat{t}_0\left(K_\infty^C,\bar{p}^W\right)} \left(P\left(K_\infty^C,t\right) + \frac{K_\infty^C}{N}P_q\left(K_\infty^C,t\right) - c\right) f(t)\,dt + \int_{\hat{t}_0(K_\infty^C,\bar{p}^W)}^{+\infty}\left(\bar{p}^W - c\right) f(t)\,dt - r \\
&= \int_{\hat{t}(K_\infty^C,c)}^{\hat{t}_0\left(K_\infty^C,\bar{p}^W\right)} \left(P\left(K_\infty^C;t\right) + \frac{K_\infty^C}{N}P_q\left(K_\infty^C;t\right) - c\right) f(t)\,dt + \int_{\hat{t}_0(K_\infty^C,\bar{p}^W)}^{+\infty}\left(\bar{p}^W - c\right) f(t)\,dt \\
&\quad - \int_{\hat{t}(K_\infty^C,c)}^{+\infty}\left(P\left(K_\infty^C;t\right) + \frac{K}{N}P_q\left(K_\infty^C;t\right) - c\right) f(t)\,dt \\
&= \int_{\hat{t}_0(K_\infty^C,\bar{p}^W)}^{+\infty}\left(\bar{p}^W - \left(P\left(K_\infty^C;t\right) + \frac{K}{N}P_q\left(K_\infty^C;t\right)\right)\right) f(t)\,dt \\
&> \int_{\hat{t}_0(K_\infty^C,\bar{p}^W)}^{+\infty}\left(\bar{p}^W - MR^\infty\right) f(t)\,dt > 0.
\end{aligned}$$

Thus, $K^C_\infty < K^C\left(\bar{p}^W\right)$, hence p^∞ is not the capacity-maximizing cap.

Necessary first-order condition The analysis above proves existence of an interior maximum, denoted $\hat{p}^W$. Since $K^C(.)$ is continuously differentiable, we must have $\frac{dK^C}{dp}\left(\hat{p}^W\right) = 0$.

C Properties of the $VoLL$

When capacity is constrained and if curtailment occurs, total differentiation of the energy balance

$$K = \alpha D\left(v\left(p^R, \gamma^*; t\right), t\right) + (1-\alpha)\mathcal{D}\left(p^R, \gamma; t\right)$$

with respect to K yields

$$\frac{\partial \gamma^*}{\partial K} = \frac{1}{\alpha \frac{\partial D}{\partial p}\frac{\partial v}{\partial \gamma} + (1-\alpha)\frac{\partial \mathcal{D}}{\partial \gamma}}.$$

$\frac{\partial v}{\partial \gamma} \leq 0$ guarantees that $\frac{\partial \gamma^*}{\partial K} > 0$ and $\frac{\partial v}{\partial K} = \frac{\partial v}{\partial \gamma}\frac{\partial \gamma^*}{\partial K} \leq 0$.

Total differentiation of the energy balance with respect to t yields

$$\frac{\partial \gamma^*}{\partial t} = -\frac{\alpha\left(\frac{\partial D}{\partial p}\frac{\partial v}{\partial t} + \frac{\partial D}{\partial t}\right) + (1-\alpha)\frac{\partial \mathcal{D}}{\partial t}}{\alpha \frac{\partial D}{\partial p}\frac{\partial v}{\partial \gamma} + (1-\alpha)\frac{\partial \mathcal{D}}{\partial \gamma}}.$$

Conditions (iii) and (iv) guarantee that the optimal serving ratio decreases as the state of the world increases: $\frac{\partial \gamma^*}{\partial t} < 0$. If curtailment occurs in state $\bar{t}$, it also occurs in all states $t \geq \bar{t}$. Furthermore, price increases as the state of the world increases:

$$\frac{dv}{dt} = \frac{\partial v}{\partial \gamma}\frac{\partial \gamma^*}{\partial t} + \frac{\partial v}{\partial t} > 0.$$

D Existence and property of Φ

Define

$$g(p) = p - v\left(p^R, 1, \bar{t}\left(K^C(p)\right)\right).$$

A root of $g(.)$ is a fixed point of $v\left(p^R, 1, \bar{t}\left(K^C(.)\right)\right)$. If $p = c + r$, the same argument as for $\alpha = 1$ shows that the price cap is binding immediately: $\hat{t}_0\left(K^C(c+r), (c+r)\right) = 0$. Then,

$$g(c+r) = c + r - v\left(p^R, 1, \bar{t}\left(K^C(c+r)\right)\right) \leq c + r - v\left(p^R, 1, 0\right) < 0$$

since $\bar{t}\left(K^C(c+r)\right) \geq 0$, $v_t > 0$, and $v\left(p^R, 1, 0\right) > (c+r)$ by assumption. Suppose $g(.)$ admits a root on $((c+r), p^\infty)$; i.e., there exists $p < p^\infty$ such that $g(p) = 0$. The maximum admissible cap is the smallest root: since $g(c+r) < 0$, $g(p) \leq 0$ for $p \leq \Phi$.

If $g(p) < 0$ for all $p < p^\infty$, the maximum admissible cap is the never binding cap; i.e., $\Phi = p^\infty$.

By definition,

$$\rho\left(K^C(\Phi), \hat{t}_0\left(K^C(\Phi), \Phi\right)\right) = \Phi,$$

and

$$\rho\left(K^C(\Phi), \bar{t}\left(K^C(\Phi)\right)\right) = v\left(p^R, 1, \bar{t}\left(K^C(\Phi)\right)\right) = \Phi.$$

Since $v\left(p^R, 1, \bar{t}\left(K^C(\Phi)\right)\right) = \Phi$,

$$\rho\left(K^C(\Phi), \hat{t}_0\left(K^C(\Phi), \Phi\right)\right) = \rho\left(K^C(\Phi), \bar{t}\left(K^C(\Phi)\right)\right) \Leftrightarrow \hat{t}_0\left(K^C(\Phi), \Phi\right) = \bar{t}\left(K^C(\Phi)\right)$$

since $\rho_t > 0$.

E Proof of Proposition 2

Since $K^C(.)$ is continuous on $[(c+r), \Phi]$, it admits a maximum $\hat{p} \in [(c+r), \Phi]$. From equation (4), $\frac{dK^C}{dp}(\Phi) < 0$ if and only if

$$1 - F\left(\hat{t}_0\right) - (\Phi - c) f\left(\hat{t}_0\right) \frac{\partial \hat{t}_0}{\partial p} < 0 \Leftrightarrow \rho_t\left(K^C(\Phi), \hat{t}_0\right) < (\Phi - c) h\left(\left(\hat{t}_0\right)\right).$$

Differentiation of equation (4) yields

$$\frac{d^2 K^C}{dp^2} = \frac{\left(-\frac{\partial \Psi}{\partial K}\right)\left(\frac{\partial^2 \Psi}{\partial K \partial p}\frac{dK^C}{dp} + \frac{\partial^2 \Psi}{\partial p^2}\right) - \frac{\partial \Psi}{\partial p}\frac{\partial^2 \Psi}{\partial K^2}}{\left(\frac{\partial \Psi}{\partial K}\right)^2},$$

hence

$$\frac{d^2K^C}{dp^2}(\hat{p}) = \frac{\frac{\partial^2\Psi}{\partial p^2}}{-\frac{\partial\Psi}{\partial K}}\left(K^C(\hat{p}), \hat{p}\right)$$

since

$$\frac{dK^C}{dp}(\hat{p}) = \frac{\partial\Psi}{\partial p}\left(K^C(\hat{p}), \hat{p}\right) = 0$$

by definition of $\hat{p}$. Then,

$$\frac{\partial^2\Psi}{\partial p^2} = -\left(2f(\hat{t}_0)\frac{\partial\hat{t}_0}{\partial p} + (p-c)\left(f'(\hat{t}_0)\left(\frac{\partial\hat{t}_0}{\partial p}\right)^2 + f(\hat{t}_0)\frac{\partial^2\hat{t}_0}{\partial p^2}\right)\right).$$

Observing that

$$\frac{\partial^2\hat{t}_0}{\partial p^2} = -\frac{\rho_{t^2}}{(\rho_t)^2}\frac{1}{\rho_t},$$

yields

$$\frac{\partial^2\Psi}{\partial p^2} = -\frac{\partial\hat{t}_0}{\partial p}f(\hat{t}_0)\left(2 + \frac{p-c}{\rho_t}\left(\frac{f'(\hat{t}_0)}{f(\hat{t}_0)} - \frac{\rho_{t^2}}{\rho_t}\right)\right).$$

Since; $\frac{p-c}{\rho_t} = \frac{1}{h(\hat{t}_0)}$ for $p = \hat{p}$,

$$\frac{\partial^2\Psi}{\partial p^2}(\hat{p}) = -\frac{\partial\hat{t}_0}{\partial p}f(\hat{t}_0)\left(2 + \frac{1}{h(\hat{t}_0)}\left(\frac{f'(\hat{t}_0)}{f(\hat{t}_0)} - \frac{\rho_{t^2}}{\rho_t}\right)(\hat{p})\right)$$

and

$$2 + \frac{1}{h(\hat{t}_0)}\left(\frac{f'(\hat{t}_0)}{f(\hat{t}_0)} - \frac{\rho_{t^2}}{\rho_t}\right)(\hat{p}) > 0 \Leftrightarrow \frac{\partial^2\Psi}{\partial p^2}(\hat{p}) > 0 \Leftrightarrow \frac{d^2K^C}{dp^2}(\hat{p}) < 0.$$

Thus, $\hat{p}$ is the unique maximum of $K^C(.)$ on $[(c+r), \Phi)$. The argument proceeds by contradiction: if there was another p^{**} such that $\frac{dK^C}{dp}(p^{**}) = 0$, then $\frac{d^2K^C}{dp^2}(p^{**}) < 0$: p^{**} is another local maximum. Since $K^C(.)$ is continuous, there would be a local minimum between these two local maxima, i.e. a price p^{***} such that $\frac{dK^C}{dp}(p^{***}) = 0$. Then $\frac{d^2K^C}{dp^2}(p^{***}) < 0$: p^{***} cannot be a local minimum. This constitutes a contradiction. Hence $K^C(.)$ is concave on $[(c+r), \Phi)$, and $\hat{p}$ is the unique maximum of $K^C(.)$ on $((c+r), \Phi)$.

Suppose now $\frac{dK^C}{dp}(\Phi) > 0$. Φ is the unique maximum of $K^C(.)$ on $[(c+r), \Phi]$. The proof proceeds again by contradiction. Suppose there exists $\hat{p} < \Phi$ such that $\frac{dK^C}{dp}(\hat{p}) = 0$. By the previous argument, $\hat{p}$ is the unique maximum, hence $K^C(.)$ is concave on $[(c+r), \Phi)$, thus $\frac{dK^C}{dp}(\Phi) < 0$, which constitutes a contradiction. Thus, $\frac{dK^C}{dp}(p) > 0$ for all $p \in [(c+r), \Phi]$, and Φ is the unique maximum of $K^C(.)$ on $[(c+r), \Phi]$.

F Proof of Proposition 3

The proof proceeds in three steps. First, I prove that $W^\infty(K)$, the expected net surplus given imperfect competition and absent any price cap, is a concave function of installed capacity with a unique maximum K^*_∞. Second, I prove that $K^C_\infty < K^*_\infty$, hence $W^\infty(K)$ is increasing for $K \leq K^C_\infty$. Finally, I prove that imposition of a price cap reduces expected net surplus. Specifically, for capacity $K^C(p)$, the expected net surplus with cap p is lower than without a cap: $W\left(K^C(p), p\right) < W^\infty\left(K^C(p)\right)$. Then, since $K^C(p) < K^C_\infty$, the result follows.

Suppose the $VoLL$ is reached on-peak. Consider first that no price cap is imposed. The net surplus off-peak is

$$\tilde{S}(t) = \alpha S\left(\rho\left(Q^C(c,t), t\right), t\right) + (1-\alpha) S\left(p^R, t\right) - cQ^C(c,t).$$

For capacity K, the net surplus on-peak before the price reaches the $VoLL$ is

$$\tilde{S}(K,t) = \alpha S(\rho(K,t), t) + (1-\alpha) S\left(p^R, t\right) - cK;$$

and is

$$\tilde{\mathcal{S}}(K,t) = \alpha S(\rho(K,t), t) + (1-\alpha) \mathcal{S}\left(p^R, \gamma^*, t\right) - cK$$

after the price reaches the $VoLL$.

The expected net surplus for capacity K is

$$W^\infty(K) = \int_0^{\hat{t}(K,c)} \tilde{S}(t) f(t)\, dt + \int_{\hat{t}(K,c)}^{\bar{t}(K)} \tilde{S}(K,t) f(t)\, dt + \int_{\bar{t}(K)}^{+\infty} \tilde{\mathcal{S}}(K,t) f(t)\, dt - rK.$$

We have:

$$\frac{dW^\infty}{dK} = \int_{\hat{t}(K,c)}^{\bar{t}(K)} \frac{\partial \tilde{S}}{\partial K}(K,t) f(t)\, dt + \int_{\bar{t}(K)}^{+\infty} \frac{\partial \tilde{\mathcal{S}}}{\partial K}(K,t) f(t)\, dt - r.$$

Then,

$$\frac{\partial \tilde{S}}{\partial K}(K,t) = \rho(K,t) - c$$

and

$$\begin{aligned}\frac{\partial\tilde{S}}{\partial K}(K,t) &= \rho(K,t)\left(1-(1-\alpha)\frac{\partial D}{\partial\gamma}\frac{\partial\gamma^*}{\partial K}\right)+(1-\alpha)\frac{\partial\mathcal{S}}{\partial\gamma}\frac{\partial\gamma^*}{\partial K}-c\\ &= \rho(K,t)+(1-\alpha)\left(\frac{\partial\mathcal{S}}{\partial\gamma}-\frac{\partial D}{\partial\gamma}\rho(K,t)\right)\frac{\partial\gamma^*}{\partial K}-c\\ &= \rho(K,t)+(1-\alpha)\left(v\left(p^R,\gamma^*,t\right)-\rho(K,t)\right)\frac{\partial D}{\partial\gamma}\frac{\partial\gamma^*}{\partial K}=\rho(K,t)-c.\end{aligned}$$

Thus,

$$\frac{dW^\infty}{dK}=\int_{\hat{t}(K,c)}^{+\infty}(\rho(K,t)-c)\,f(t)\,dt-r,$$

hence

$$\frac{d^2W^\infty}{dK^2}=\int_{\hat{t}(K,c)}^{+\infty}\rho_q(K,t)\,f(t)\,dt-\left(\rho\left(\hat{t}(K,c),t\right)-c\right)f\left(\hat{t}\right)\frac{\partial\hat{t}}{\partial K}<0.$$

$W^\infty(.)$ is globally concave. The usual arguments show that there exists a unique K^*_∞ that maximizes $W^\infty(.)$, uniquely defined by

$$\frac{dW^\infty}{dK}(K^*_\infty)=\int_{\hat{t}(K^*_\infty,c)}^{+\infty}(\rho(K^*_\infty,t)-c)\,f(t)\,dt-r=0.$$

One immediately verifies that

$$\frac{dW^\infty}{dK}\left(K^C_\infty\right)=\int_{\hat{t}(K^C_\infty,c)}^{+\infty}\left(\rho\left(K^C_\infty,t\right)-c\right)f(t)\,dt-r=\frac{K^C_\infty}{N}\int_{\hat{t}(K^C_\infty,c)}^{+\infty}\left(-\rho_q\left(K^C_\infty,t\right)\right)f(t)\,dt>0,$$

thus

$$K^C_\infty<K^*_\infty\Rightarrow W^\infty\left(K^C_\infty\right)<W^\infty(K^*_\infty).$$

Consider now a price cap p is imposed, which is reached on-peak. The net surplus after the price reaches the $VoLL$ is

$$\mathring{\mathcal{S}}(K,t)=\alpha S(p,t)+(1-\alpha)\,\mathcal{S}\left(p^R,\mathring{\gamma},t\right)-cK$$

where $\mathring{\gamma}$ such that

$$\alpha D(p,t)+(1-\alpha)\,\mathcal{D}\left(p^R,\mathring{\gamma},t\right)=K.$$

Observe that

$$\mathring{\mathcal{S}}(K,t) < \tilde{\mathcal{S}}(K,t)$$

since price reactive consumers do not face the opportunity cost of power. Thus the expected net surplus is

$$\begin{aligned} W_1\left(K^C(p),p\right) &= \int_0^{\hat{t}\left(K^C(p),c\right)} \tilde{S}(t)\, f(t)\, dt + \int_{\hat{t}(K^C(p),c)}^{\hat{t}_0\left(K^C(p),p\right)} \tilde{S}\left(K^C(p),t\right) f(t)\, dt \\ &\quad + \int_{\hat{t}_0(K^C(p),p)}^{+\infty} \mathring{\mathcal{S}}(K,t)\, f(t)\, dt - rK^C(p) \end{aligned}$$

Therefore,

$$\begin{aligned} W_1\left(K^C(p),p\right) &= W^\infty\left(K^C(p)\right) - \int_{\hat{t}_0(K^C(p),p)}^{\bar{t}\left(K^C(p)\right)} \tilde{S}(K,t)\, f(t)\, dt - \int_{\bar{t}(K^C(p))}^{+\infty} \tilde{\mathcal{S}}(K,t)\, f(t)\, dt \\ &\quad + \int_{\hat{t}_0(K^C(p),p)}^{+\infty} \mathring{\mathcal{S}}(K,t)\, f(t)\, dt \\ &= W^\infty\left(K^C(p)\right) - \int_{\hat{t}_0(K^C(p),p)}^{\bar{t}\left(K^C(p)\right)} \left(\tilde{S}(K,t) - \mathring{\mathcal{S}}(K,t)\right) f(t)\, dt \\ &\quad - \int_{\bar{t}(K^C(p))}^{+\infty} \left(\tilde{\mathcal{S}}(K,t) - \mathring{\mathcal{S}}(K,t)\right) f(t)\, dt \\ &< W^\infty\left(K^C(p)\right) \end{aligned}$$

since $\tilde{S}(K,t) - \mathring{\mathcal{S}}(K,t) > 0$ and $\tilde{\mathcal{S}}(K,t) - \mathring{\mathcal{S}}(K,t) > 0$. For the same capacity, imposing a price cap reduces net surplus. Then, since $K^C(p) < K^C_\infty$,

$$W_1\left(K^C(p),\Phi\right) < W^\infty\left(K^C(p)\right) < W^\infty\left(K^C_\infty\right),$$

which completes the proof.

Similar arguments apply if the cap or the $VoLL$ are reached off-peak.

Technical Report

Not part of the paper

Available upon request

G Derivation of the Cournot capacity

For the reader's convenience, the derivation of the Cournot equilibrium capacity is presented here. The proof follows Zöttl (2011). The main difference is the introduction of rationing.

G.1 Capacity constrained Cournot equilibrium without a price cap

Producers are ordered by increasing capacity: $k^1 \leq k^2 ... \leq k^N$.

The unconstrained Cournot aggregate output $Q^C(t)$ in state t is defined by $\rho(Q;t) - c + \frac{Q}{N}\rho_q(Q;t) = 0$. Under Assumption 4, the implicit function theorem yields:

$$\frac{dQ^C}{dt} = -\left(\rho_t + \frac{Q}{N}\rho_{qt}\right)\left(\rho_q + \frac{\rho_q + Q\rho_{qq}}{N}\right)^{-1} > 0$$

Lemma 2 *Define $t^0 = 0$. For a given vector $\mathbf{k} \in \mathbb{R}^N$ of generation capacities, there exists N critical states of the world $0 \leq t^1 \leq t^2 \leq ... \leq t^N$ such that $\hat{q}^n(\mathbf{k};t)$, the equilibrium output for producer n, is characterized by*

$$\hat{q}^n(\mathbf{k};t) = \begin{cases} \phi^{j+1}(k^1, ..., k^j; t) & if \quad t \in [t^j, t^{j+1}] \;\; for \; j < n \\ k^n & if \quad\quad t \geq t^n \end{cases}$$

where $\phi^{j+1}(k^1, ..., k^j; t)$ defined on $[t^j, t^{j+1}]$ is the solution of a "modified Cournot condition":

$$\rho\left(\sum_{i=1}^{j} k^i + (N-j)\phi^{j+1}; t\right) - c + \phi^{j+1}\rho_q\left(\sum_{i=1}^{j} k^i + (N-j)\phi^{j+1}; t\right) = 0$$

$\forall n \leq N$, $\hat{q}^n(\mathbf{k};t)$ is continuous in all its arguments and increasing in t.

Producers expected profit is:

$$\Pi^n(k^n; \mathbf{k}_{-n}) = \sum_{j=0}^{n-1} \int_{t^j}^{t^{j+1}} \phi^{j+1}\left(\rho\left(\hat{Q}\right) - c\right) f(t)\, dt + k^n \sum_{j=n}^{N} \left[\int_{t^j}^{t^{j+1}} \left(\rho\left(\hat{Q}\right) - c\right) f(t)\, dt - r\right]$$

where $\hat{Q}(\mathbf{k};t) = \sum_{n=1}^{N} \hat{q}^n(\mathbf{k};t)$ is the aggregate output.

Proof. *Construction of the equilibrium proceeds by induction on $n \leq N$. As seen previously, $Q^C(t)$ is increasing in t. Denote t^1 the first state such that $Q^C(t^1) = k^1$. Suppose $t^1 \to +\infty$, then $\forall t \geq 0$,*

$Q^C(t) < k^1 \leq k^2 \leq ... \leq k^N$. Then:

$$\Pi^1(k^1; \mathbf{k}_{-1}) = \mathbb{E}\left[Q^C(t)\left(\rho\left(Q^C(t); t\right) - c\right)\right] - rk^1$$

and $\frac{\partial \Pi^1}{\partial k^1} = -r < 0$, hence $\Pi^1\left(k^1\right) < 0$ since $\Pi^1(0) = 0$. This contradicts producer 1 individual rationality, hence t^1 exists by contradiction. Denote $\hat{Q}^1(t) = Q^C(t)$ and $\phi^1(t) = \frac{Q^C(t)}{N}$.

Suppose now we have characterized the equilibrium up until t^n, defined by $\phi^n(k^1, ..., k^{n-1}; t^n) = k^n$. We now search for an equilibrium for $t \geq t^n$. Suppose all generators $i = 1, .., n$ produce up to their capacity k^i. The profit of any generator $j > n$ is:

$$\pi^j(q^j; \mathbf{q}_{-j}; t) = q^j\left(\rho\left(\sum_{i=1}^{n} k^i + q^j + \sum_{\substack{i=n+1 \\ i \neq j}}^{N} q^i; t\right) - c\right)$$

As long as producer $(n+1)$ is not constrained, we have:

$$\frac{\partial \pi^j}{\partial q^j} = \rho\left(\sum_{i=1}^{n} k^i + q^j + \sum_{\substack{i=n+1 \\ i \neq j}}^{N} q^i; t\right) - c + q^j \rho_q\left(\sum_{i=1}^{n} k^i + q^j + \sum_{\substack{i=n+1 \\ i \neq j}}^{N} q^i; t\right)$$

Since the first-order conditions are symmetric, a symmetric interior equilibrium $\phi^{n+1}(k^1, ..., k^n; t)$ is characterized by:

$$\rho\left(\sum_{i=1}^{n} k^i + (N-n)\phi^{n+1}; t\right) - c + \phi^{n+1}\rho_q\left(\sum_{i=1}^{n} k^i + (N-n)\phi^{n+1}; t\right) = 0$$

Since $\pi^j(.; t)$ is strictly concave, ϕ^{n+1} is a maximum, hence it constitutes a best response to the others' strategies. ϕ^{n+1} is increasing in t, since by the implicit function theorem:

$$\frac{\partial \phi^{n+1}}{\partial t} = -\frac{\rho_t + \phi^{n+1}\rho_{qt}}{(N-n+1)\left(\rho_q + \frac{N-n}{N-n+1}\phi^{n+1}\rho_{qq}\right)} > 0$$

Furthermore, $\phi^{n+1}(t^n) = \phi^n(t^n) = k^n$. To see that, we observe that $\phi^{n+1}(t^n)$ verifies:

$$\frac{\partial \pi^j}{\partial q^j}(t^n) = \rho\left(\sum_{i=1}^{n} k^i + (N-n)\phi^{n+1}; t^n\right) - c + \phi^{n+1}\rho_q\left(\sum_{i=1}^{n} k^i + (N-n)\phi^{n+1}; t^n\right) = 0$$

while $\phi^n(t)$ *solves:*

$$\rho\left(\sum_{i=1}^{n-1} k^i + (N-n+1)\,\phi^n; t\right) - c + \phi^n \rho_q \left(\sum_{i=1}^{n-1} k^i + (N-n+1)\,\phi^n; t\right) = 0$$

hence, since $\hat{q}^n(t^n) = k^n$ *by construction:*

$$\rho\left(\sum_{i=1}^{n} k^i + (N-n)\,k^n; t^n\right) - c + k_n \rho_q \left(\sum_{i=1}^{n} k^i + (N-n)\,k^n; t^n\right) = 0$$

Since $\pi^j(.;t^n)$ *is strictly concave, there exists a unique value such that* $\frac{\partial \pi^j}{\partial q^j}(t^n) = 0$ *hence* $\phi^{n+1}(t^n) = k^n = \phi^n(t^n)$.

Producer $(n+1)$ *produces* $\phi^{n+1}(k^1, ..., k^n; t)$ *up to* t^{n+1} *defined by* $\phi^{n+1}(k^1, ..., k^n; t) = k^{n+1}$. *As before, we can show by contradiction that* t^{n+1} *exists. Furthermore, since* $k^{n+1} \geq k^n$*, then* $t^{n+1} \geq t^n$.

We now show that $q^l = k^l$ *for* $l \leq n$ *is a best response to* $q^j = \phi^{n+1}$ *for* $j > n$. *Suppose* $q^i = k^i\ \forall i \leq n, i \neq l$ *and* $q^j = \phi^{n+1}$ *for* $j > n$. *Then:*

$$\frac{\partial \pi^l}{\partial q^l} = \rho\left(\sum_{\substack{i=1 \\ i \neq j}}^{n} k^i + q^l + (N-n)\,\phi^{n+1}; t\right) - c + q^l \rho_q \left(\sum_{\substack{i=1 \\ i \neq j}}^{n} k^i + q^l + (N-n)\,\phi^{n+1}; t\right)$$

hence

$$\begin{aligned} \left.\frac{\partial \pi^l}{\partial q^l}\right|_{q^l = k^l} &= \rho\left(\sum_{i=1}^{n} k^i + (N-n)\,\phi^{n+1}; t\right) - c + k^l \rho_q \left(\sum_{i=1}^{n} k^i + (N-n)\,\phi^{n+1}; t\right) \\ &= -\left(\phi^{n+1} - k^l\right) \rho_q \left(\sum_{i=1}^{n} k^i + (N-n)\,\phi^{n+1}; t\right) > 0 \end{aligned}$$

since $\phi^{n+1}(t) > \phi^{n+1}(t^n) = k^n \geq k^j$ *for* $t > t^n$.

We have therefore completed step $(n+1)$. *By induction, the structure of the equilibria holds up until* $n = N$*, as long as we adopt the convention:* $t^{N+1} \to +\infty$, $\phi^{N+1}(t) = 0$ *and* $\hat{Q}(t^N) = K = \hat{Q}(t)$ $\forall t \geq t^N$.

From the previous discussion, $\hat{q}^n(\mathbf{k};t)$ *and* $\hat{Q}(\mathbf{k};t)$ *are continuous and increasing in* t.

The expression of profits follow directly from the characterization of equilibria above. ■

G.2 Capacity constrained Cournot equilibrium with price cap

Lemma 3 *Suppose the capacity constraint binds before the price cap constraint, i.e.,* $t^N \leq \bar{t}\left(K, \bar{p}^W\right)$ *at the equilibrium. Producer n's equilibrium profit for the constrained Cournot game in state t is:*

$$\begin{aligned} \Pi^n\left(k^n;\mathbf{k}_{-n}\right) &= \sum_{j=0}^{n-1}\int_{t^j}^{t^{j+1}} \phi^{j+1}\left(\rho\left(\hat{Q}\right)-c\right) f\left(t\right)dt \\ &+k^n\left\{\sum_{j=n}^{N}\left[\int_{t^j}^{t^{j+1}}\left(\rho\left(\hat{Q}\right)-c\right) f\left(t\right)dt\right]+\int_{\bar{t}\left(K,\bar{p}^W\right)}^{+\infty}\left(\bar{p}^W-c\right) f\left(t\right)dt-r\right\} \end{aligned} \tag{18}$$

where t^j *for* $j = 1, ..., N$ *is the first state of the world such that all producers up to* j *are capacity constrained,* $t^0 = 0$ *and the convention* $t^{N+1} = \bar{t}\left(K, \bar{p}^W\right)$, ϕ^{j+1} *is producer n's output in state* $t \in \left[t^j, t^{j+1}\right]$ *for* $j < n$, *and* $\hat{Q}$ *is the equilibrium aggregate output.*

Proof. *In the off-peak states of the world where at least one generator is unconstrained, i.e., with our previous notation* $t < t^N(K) \leq \bar{t}\left(K, \bar{p}^W\right)$, *imposition of the price cap has no impact on the equilibrium in these states, and Lemma 2 applies.*

Consider now the peak states of the world $t \geq t^N(K)$, *and* $\hat{Q}(t) = K = \sum_{m=1}^{N} k^m$. *As long as* $\rho(K;t) \leq \bar{p}^W$, *imposition of the price cap has no impact on the equilibrium in these states, and Lemma 2 applies.*

States of the world may exist where $\rho(K;t) > \bar{p}^W$. *Then, for* $t \geq \bar{t}\left(K, \bar{p}^W\right)$, $p(t) = \bar{p}^W$. *Facing a constant price, generators individually maximize production to maximize profit, hence* $q^n(t) = k^n$ *for all n is an equilibrium. This then yields equation (18). However, the SO must ration demand.* ■

G.3 Equilibrium investment

Lemma 4 *For any* $(k^n, \mathbf{k}_{-n})$*:*

$$\frac{\partial \Pi^n}{\partial k^n}\left(k^n,\mathbf{k}_{-n}\right)=\sum_{j=n}^{N}\left[\int_{t^j}^{t^{j+1}}\left(\rho\left(\hat{Q}\right)-c+k^n\rho_q\left(\hat{Q}\right)\frac{\partial \hat{Q}}{\partial k^n}\right) f\left(t\right)dt\right]+\int_{\bar{t}\left(K,\bar{p}^W\right)}^{+\infty}\left(\bar{p}^W-c\right) f\left(t\right)dt-r. \tag{19}$$

For any $k^N \geq \frac{K}{N}$

$$\frac{\partial^2 \Pi^N}{(\partial k^N)^2}\left(\frac{K}{N},...,\frac{K}{N},k^N\right) = \int_{t^N}^{\bar{t}\left(K,\bar{p}^W\right)} \left[2\rho_q\left(\hat{K};t\right) + k^N \rho_{qq}\left(\hat{K};t\right)\right] f(t)\,dt + k^N \rho_q\left(\hat{K};\bar{t}\left(\hat{K},\bar{p}^W\right)\right) f\left(\bar{t}\left(\hat{K},\bar{p}^W\right)\right) \frac{\partial \bar{t}\left(\hat{K},\bar{p}^W\right)}{\partial k^N} < 0 \tag{20}$$

where $\hat{K} = k^n + \frac{N-1}{N}K$*. Furthermore,* $\Pi^n\left(\frac{K}{N},..,\frac{K}{N}\right)$ *is globally concave.*

Proof. *The first-order derivative of profit function is:*

$$\frac{\partial \Pi^n}{\partial k^n} = \sum_{j=n}^{N}\left[\int_{t^j}^{t^{j+1}}\left(\rho\left(\hat{Q}\right) - c + k^n \rho_q\left(\hat{Q}\right)\frac{\partial \hat{Q}}{\partial k^n}\right) f(t)\,dt\right] + \int_{\bar{t}(K,\bar{p}^W)}^{+\infty}\left(\bar{p}^W - c\right) f(t)\,dt - r + \Delta_1$$

where

$$\begin{aligned}\Delta_1 &= k_n \sum_{j=n}^{N}\left[\left(\rho\left(\hat{Q}\left(t^{j+1}\right);t^{j+1}\right) - c\right) f\left(t^{j+1}\right)\frac{\partial t^{j+1}}{\partial k^n} - \left(\rho\left(\hat{Q}\left(t^j\right);t^j\right) - c\right) f\left(t^j\right)\frac{\partial t^j}{\partial k^n}\right] \\ &\quad + \hat{q}^n\left(t^n\right)\left(\rho\left(\hat{Q}\left(t^n\right);t^n\right) - c\right) f\left(t^n\right)\frac{\partial t^n}{\partial k^n} - k^n\left(\bar{p}^W - c\right) f\left(\bar{t}\left(K,\bar{p}^W\right)\right)\frac{\partial \bar{t}\left(K,\bar{p}^W\right)}{\partial k^n} \\ &= 0\end{aligned}$$

since $\hat{q}^n\left(t^n\right) = k^n$ *and* $\rho\left(\hat{Q}\left(\bar{t}\left(K,\bar{p}^W\right)\right);\bar{t}\left(K,\bar{p}^W\right)\right) = \bar{p}^W$*. This proves equation (19).*

Suppose $k^N \geq \frac{K^C}{N}$ *while* $k^n = \frac{K^C}{N}$ *for all* $n < N$*. Equation (19) yields:*

$$\frac{\partial \Pi^N}{\partial k^N}\left(\frac{K^C}{N},...,\frac{K^C}{N},k^N\right) = \int_{t^N(K)}^{\bar{t}\left(K,\bar{p}^W\right)(K)}\left(\rho(K;t) + k^N\rho_q(K;t) - c\right) f(t)\,dt + \int_{\bar{t}(K,\bar{p}^W)(K)}^{+\infty}\left(\bar{p}^W - c\right) f(t)\,dt - r.$$

Thus:

$$\begin{aligned}\frac{\partial^2 \Pi^N}{\partial\left(k^N\right)^2}\left(\frac{K^C}{N},...,\frac{K^C}{N},k^N\right) &= \int_{t^N}^{\bar{t}\left(K,\bar{p}^W\right)}\left[2\rho_q\left(\hat{K};t\right) + k^N\rho_{qq}\left(\hat{K};t\right)\right] f(t)\,dt \\ &\quad + k^N\rho_q\left(\hat{K};\bar{t}\left(K,\bar{p}^W\right)\right) f\left(\bar{t}\left(K,\bar{p}^W\right)\right)\frac{\partial \bar{t}\left(K,\bar{p}^W\right)}{\partial k^N} \\ &< 0.\end{aligned}$$

This proves equation (20). Then, selecting $k^N = \frac{K}{N}$ proves the global concavity of $\Pi^n\left(\frac{K}{N}, .., \frac{K}{N}\right)$. ■

Lemma 5 *K^C solution of*

$$\int_{t^N(K^C)}^{\bar{t}(K,\bar{p}^W)(K^C)} \left(\rho\left(K^C;t\right) + \frac{K^C}{N}\rho_q\left(K^C;t\right) - c\right) f(t)\,dt + \int_{\bar{t}(K^C,\bar{p}^W)(K^C)}^{+\infty} \left(\bar{p}^W - c\right) f(t)\,dt = r$$

is the only symmetric equilibrium investment.

Proof. *If $k^n = \frac{K}{N}$, for all n, all producers are constrained simultaneously: $t^n = t^N$ for all n. The first order derivative (19) then becomes:*

$$\frac{\partial \Pi^n}{\partial k^n}\left(\frac{K}{N}, ..., \frac{K}{N}\right) = \int_{t^N(K)}^{\bar{t}(K,\bar{p}^W)(K)} \left(\rho(K;t) + \frac{K}{N}\rho_q(K;t) - c\right) f(t)\,dt + \int_{\bar{t}(K,\bar{p}^W)(K)}^{+\infty} \left(\bar{p}^W - c\right) f(t)\,dt - r.$$

$\frac{\partial \Pi^n}{\partial k^n}(0, ..., 0) = \int_0^{\bar{t}(K,\bar{p}^W)} (\rho(0;t) - c) f(t)\,dt + \int_{\bar{t}(K,\bar{p}^W)}^{+\infty} \left(\bar{p}^W - c\right) f(t)\,dt - r > \bar{p}^W - (c+r) > 0$ *since $\rho(0;t) > \bar{p}^W > (c+r)$ by equation (**??**). $\lim_{K\to+\infty} \frac{\partial \Pi^n}{\partial k^n}\left(\frac{K}{N}, ..., \frac{K}{N}\right) = -r < 0$. Hence $K^C > 0$ such that $\frac{\partial \Pi^n}{\partial k^n}\left(\frac{K^C}{N}, ..., \frac{K^C}{N}\right) = 0$ exists.*

We now prove that $\left(\frac{K^C}{N}, ..., \frac{K^C}{N}\right)$ is an equilibrium. Consider first a negative deviation: $k^1 \leq \frac{K^C}{N}$ while $k^n = \frac{K^C}{N}$ for all $n > 1$. Total installed capacity is $K = k^1 + \frac{N-1}{N}K^C \leq K^C$.

$$\begin{aligned}
\frac{\partial \Pi^1}{\partial k^1}\left(k^1, \frac{K^C}{N}, ..., \frac{K^C}{N}\right) &= \int_{t^1}^{t^N(K)} \left(\rho\left(\hat{Q}\right) + k^1\rho_q\left(\hat{Q}\right)\frac{\partial \hat{Q}}{\partial k^1} - c\right) f(t)\,dt \\
&+ \int_{t^N(K)}^{\bar{t}(K,\bar{p}^W)(K)} \left(\rho(K) + k^1\rho_q(K) - c\right) f(t)\,dt \\
&+ \int_{\bar{t}(K,\bar{p}^W)(K)}^{+\infty} \left(\bar{p}^W - c\right) f(t)\,dt - r \\
&= \int_{t^1}^{t^N(K)} \left(\rho\left(\hat{Q}\right) + k^1\rho_q\left(\hat{Q}\right)\frac{\partial \hat{Q}}{\partial k^1} - c\right) f(t)\,dt \\
&+ \int_{t^N(K)}^{\bar{t}(K,\bar{p}^W)} \left(\rho(K) + k^1\rho_q(K) - c\right) f(t)\,dt \\
&- \int_{t^N(K^C)}^{\bar{t}(K^C,\bar{p}^W)} \left(\rho\left(K^C\right) + \frac{K^C}{N}\rho_q\left(K^C\right) - c\right) f(t)\,dt \\
&+ \int_{\bar{t}(K,\bar{p}^W)}^{+\infty} \left(\bar{p}^W - c\right) f(t)\,dt - \int_{\bar{t}(K^C,\bar{p}^W)}^{+\infty} \left(\bar{p}^W - c\right) f(t)\,dt
\end{aligned}$$

$t^N(K) \leq t^N\left(K^C\right)$ and $\bar{t}\left(K,\bar{p}^W\right)(K) \leq t^{\bar{p}^W}\left(K^C\right)$ since $K < K^C$. Then:

$$\begin{aligned}\frac{\partial \Pi^1}{\partial k^1}\left(k^1,\frac{K^C}{N},...,\frac{K^C}{N}\right) &= \int_{t^1}^{t^N(K)}\left(\rho\left(\hat{Q}\right)+k^1\rho_q\left(\hat{Q}\right)\frac{\partial \hat{Q}}{\partial k^1}-c\right)f(t)\,dt \\ &+\int_{t^N(K)}^{t^N\left(K^C\right)}\left(\rho(K)+k^1\rho_q(K)-c\right)f(t)\,dt \\ &+\int_{t^N(K)}^{\bar{t}\left(K,\bar{p}^W\right)}\left(\rho(K)+k^1\rho_q(K)-\left(\rho\left(K^C\right)+\frac{K^C}{N}\rho_q\left(K^C\right)\right)\right)f(t)\,dt \\ &+\int_{\bar{t}\left(K,\bar{p}^W\right)}^{\bar{t}\left(K^C,\bar{p}^W\right)}\left(\bar{p}^W-\rho\left(K^C\right)-\frac{K^C}{N}\rho_q\left(K^C\right)\right)f(t)\,dt.\end{aligned}$$

$\rho\left(\hat{Q}\right)+k^1\rho_q\left(\hat{Q}\right)-c=\left(k^1-\phi^N\right)\rho_q\left(\hat{Q}\right)\frac{\partial\hat{Q}}{\partial k^1}\geq 0$ for $t\in\left[t^1,t^N(K)\right]$. $\rho\left(K;t^N(K)\right)+k^1\rho_q\left(K;t^N(K)\right)=c$, and $\rho_t(K)+k^1\rho_{qt}(K)\geq 0$, hence $\rho(K)+k^1\rho_q(K)-c\geq 0$ for $t\in\left[t^N(K),t^N\left(K^C\right)\right]$. $\rho_q(Q)+q\rho_{qq}(Q)<0$, hence $\rho(K)+k^1\rho_q(K)\geq\rho(K)+\frac{K^C}{N}\rho_q(K)\geq\rho\left(K^C\right)+\frac{K^C}{N}\rho_q\left(K^C\right)$ $\left(k^1-\frac{K^C}{N}\right)\rho_q\left(\hat{Q}\right)\frac{\partial\hat{Q}}{\partial k^1}\geq 0$ for $t\in\left[t^N(K),\bar{t}\left(K,\bar{p}^W\right)(K)\right]$. Finally, $\rho\left(K^C\right)\leq\bar{p}^W$ for $t\leq t^{\bar{p}^W\left(K^C\right)}$, hence $\bar{p}^W-\rho\left(K^C\right)-\frac{K^C}{N}\rho_q\left(K^C\right)\geq 0$ for $t\in\left[\bar{t}\left(K,\bar{p}^W\right)(K),t^{\bar{p}^W\left(K^C\right)}\right]$. Thus, $\frac{\partial\Pi^1}{\partial k^1}\left(k^1,\frac{K^C}{N},...,\frac{K^C}{N}\right)\geq 0$: a negative deviation is not profitable.

From equation (20), $\frac{\partial\Pi^N}{\partial k^N}\left(\frac{K^C}{N},...,\frac{K^C}{N},k^N\right)$ is decreasing for $k^N\geq\frac{K^C}{N}$, hence $\frac{\partial\Pi^N}{\partial k^N}\left(\frac{K^C}{N},...,\frac{K^C}{N},k^N\right)\leq 0$: a positive deviation is not profitable.

Therefore, $\left(\frac{K^C}{N},...,\frac{K^C}{N}\right)$ is a symmetric equilibrium. Furthermore, K^C is the only symmetric equilibrium since $\Pi^n\left(\frac{K}{N},...,\frac{K}{N}\right)$ is globally concave. ■