When electricity provision becomes unreliable (preliminary and incomplete)*

Catherine Bobtcheff[†] Philippe De Donder[‡] François Salanié[§]

June 8, 2022

Abstract

We set up a model of electricity provision in which delivery to consumer is only imperfectly reliable, depending on the difference between stochastic capacity and stochastic demand. We study the impact on optimal regulations of the possibility of blackouts: how the optimal price cap relates to the Value of Lost Load, how to subsidize capacity investments depending on intermittency, and whether retail electricity should be taxed. In general, answers to these questions depend on correlations between available capacity, the VOLL, and the elasticity of demand, conditional to the occurrence of a blackout.

Keywords: electricity, reliability, renewables, climate change.

JEL Classification: D24, Q41, Q42, Q48

*We thank Stefan Ambec, Claude Crampes, and various experts from the energy sector, for their insightful suggestions and remarks. This research has benefitted from the financial support of Engie in the context of the research foundation TSE-Partnership. Financial support from the ANR (Programme d'Investissement d'Avenir ANR-17-EURE-0001 and ANR-17-EURE-0010) is gratefully acknowledged.

[†]Paris School of Economics and CNRS. E-mail: catherine.bobtcheff@psemail.eu.

[‡]Toulouse School of Economics, University of Toulouse Capitole, CNRS, Toulouse, France. E-mail: philippe.dedonder@tse-fr.eu.

[§]Toulouse School of Economics, University of Toulouse Capitole, INRAE, Toulouse, France. E-mail: francois.salanie@tse-fr.eu.

Introduction

Climate change, and the policies undertaken to mitigate it, will deeply transform the energy sector. So far, the replacement of energy sources based on fossil fuels has led to the development of so-called intermittent sources that depend on the presence of light and wind, and produce electricity. Other sources in development, such as hydrogen combustion, also require electricity in the first place for the production of hydrogen. The consequence is that electricity will probably be needed for almost every use, from electric car to heating to cooling, and also for many industrial processes.

The provision of electricity will therefore require huge investments, both in production units and in the grid that delivers it to consumers. This paper argues that this costly transformation should also be analyzed in terms of risk and reliability. An electricity grid operates under strong technical constraints: in particular, supply and demand have to be precisely balanced at every date. This balance is difficult to manage when a significant part of supply is intermittent; demand depends also on weather, and cannot be managed in real-time; and additional pressure is put on production units and on the grid because climate change may foster extreme climatic events. Reliability is thus likely to become a major concern in the future, both for households and for many industrial or commercial processes. Regulation will have to take into account the probability of major blackouts. In the absence of alternative energy supply, consumers will face a risk of disconnection, with possibly dire consequences if for example heating is impossible, because electric lines are cut due a major snow storm. The recent crisis in Texas illustrates perfectly the costs and hazards of putting all eggs in one basket.

This paper aims at analyzing these risks, and how the regulation of electricity markets has to be modified if such transforms take place. Our model of the electricity sector is inspired by the models in Joskow and Tirole (2007) and Fabra (2018). Our model allows for heterogeneity of both consumers and producers, incorporates arbitrary shocks and intermittency to both capacity and demand, and proposes a general representation of disconnections, and of rolling and systemic blackouts. We compare optima to competitive equilibria, and we obtain explicit formulas for the optimal values of regulatory instruments.

The starting point of our analysis is the recognition that the main difficulties with the electricity sector lie on the demand side, and in particular on how demand is managed. One constraint is technological. At the most local level, each consumer enjoys electricity by connecting devices to his own private loop, which is itself connected to the grid. By switching on and off these devices, the consumer determines his electricity demand. For technical reasons, this demand is either fully satisfied, or not at all: partial rationing of a consumer's demand is impossible.[1] The other constraint is behavioral: because attention is costly, the typical consumer generally does not gather information on the wholesale price, and does not react to variations of this price.[2]

These two constraints severely limit the set of possible retail contracts. Strikingly, retailers all over the world have ended up selling the same type of contract: in exchange of a fixed fee, this contract allows the consumer to demand the quantity he wishes (maybe up to a pre-specified ceiling) at a pre-specified retail price. In the following, and for simplicity, we consider that all consumers sign this type of contract.

The use of such contracts puts retailers in an awkward situation: they sell at a fixed price a good whose wholesale price is varying in real time. This in turn creates two difficulties. Firstly, an adverse selection problem appears, as the profits made on each consumer depend on his load profile, i.e., on whether he consumes electricity at peak hours or not (Joskow and Tirole, 2006). If this load profile is private information, then competition on the retail market is likely to lead to inefficient allocations. Once more, for the sake of simplicity we shall ignore this problem, for example by assuming that the consumer's type is public information. Secondly, under some contingencies it becomes possible that demand lies above available capacity: then even a very high wholesale price would not balance supply and demand, since consumers do not react to changes in the wholesale price, and the capacity constraint is binding anyway. And one indeed expects this price to take very high values, since every producer is indispensable.

[1]This would require to set priorities at the level of each device, or to install several distinct loops that could be switched on and off independently. In the future, one indeed expects an increase in the use of smart devices able to manage their activity as a function of an information on the real-time wholesale price of electricity. This article ignores this possibility.

[2]Once more, this is not the case for many industrial users. And once more, we choose to ignore this possibility in this article, for the sake of simplicity.

Regulators typically reacts to these crisis by rationing consumers, and by applying a pre-specified price cap. The former intervention is the only way to balance the market. Interestingly, regulators typically choose to disconnect the minimum number of consumers consistent with the equality between demand and available capacity, and by doing so they accept to perpetuate the market power of producers since all of them remain indispensable; we stick to this case in our model.[3] The second intervention then becomes compulsory, since in the absence of a price cap the wholesale price would still spiral upwards without limits. It is thus not only motivated by the need to settle transactions between producers and retailers at some reasonable price, but also by other objectives such as the protection of retailers from bankruptcy, or the mitigation of the producers' market power (see Wolak, 2013, among others).

Setting a price cap implies two potential distortions. On the retail side, retailers benefit from a lower price of electricity, and therefore the competitive retail price is reduced: this may justify the use of a tax on retail electricity. On the supply side, in case of crisis the producers' payoffs are reduced, and this makes investments in capacity less profitable. How to deal with this "'missing money" problem is the topic of many works in the literature, and we simply deal with it by introducing the possibility of a subsidy to capacity.

Our model then allows to define optima, equilibria, and how to decentralize the former allocations using the above instruments: a price cap in case of blackouts, a retail tax, and subsidies to capacity investments. Here are a few questions which will be answered soon (TBC!):

• When is it the case that an optimally set price cap is sufficient for decentralizing optimal allocations?

Answer: when there is no intermittency, and demand has a fixed elasticity, whatever the state of nature. Then the price cap should be set to the expectation of the Value of Lost Load, conditional on the occurrence of a blackout.

[3]Regulators display a willingness to avoid orderly blackouts which seems sometimes out of proportion with standard estimates of the cost of these episodes; on this point, see Wolak (2013) and Joskow (2021, pp. 5-6). In a simple model of demand rationing with responsive and non-responsive consumers, Gerlagh, Liski, and Vehviäinen (2022) relaxes this assumption, and makes both the level of rationing and the price cap contingent on market information. We leave an analysis of this possibility for future research.

• This means that the optimal price cap takes very high values. For reasons we have already listed, regulators tend to set much lower price caps. Is decentralization still possible?

Answer: yes, but both a retail tax and subsidies to capacity are needed.

• When is it the case that subsidies to capacity should be uniform?

Answer: under the same conditions as for the first question: when there is no intermittency (so that a unit of capacity is well-defined, whatever the type of production unit), and demand has a fixed elasticity.

• Under intermittency, how should we compute optimal capacity subsidies?

Answer: they depend on complex correlations between several variables: the available capacity, the VOLL, and the elasticity of demand, conditional to the occurrence of a blackout.

• In project: whether alternative sources of energy (such as biogas or hydrogen) may represent useful substitutes during these crisis.

1 The Model

State of nature: We begin by defining an exogenous shock $s \in S$, with known distribution. This shock captures the effects on both supply and demand of weather, earthquakes, fires, breakdowns, and of all elements considered as exogenous, i.e., beyond the control of agents.

Timing: Ex-ante, a regulatory policy is set, and electricity producers invest in production units, while retailers offer retail contracts to consumers. Then the state of nature s is realized, and it is observed by all agents. Ex-post, consumers choose their demand, producers decide what to produce, and the wholesale price is determined on the market, depending on stochastic blackouts. Retailers pay the wholesale price to producers and distribute electricity to consumers, or at least to those producers and consumers that are still connected to the grid.

Supply: Electricity is produced by different types of units, indexed by the subscript k. A unit of type k has a variable cost function $c_k(q_k, s)$ in state s, weakly convex in production $q_k \in [0, K_k(s)]$. Note that the upper bound is a finite production capacity, which we allow to depend on the state s, so as to capture for example the dependence of intermittent production units on weather.

Ex-ante, it is possible to invest in an arbitrary number $x_k \geq 0$ of units of type k, at an investment cost $I_k(x_k)$, assumed convex. We avoid indivisibilities by allowing x_k to take non-integer values. For the sake of simplicity, we also specialize the analysis to the case when all numbers x_k are chosen to be strictly positive: this will avoid lengthy discussions of corner solutions. Finally, note that the investment cost is allowed to vary non-linearly with the number of units, so as to capture the idea that for some types of units, such as dams or wind farms, there is a limited number of possible locations with the same characteristics.

Given the vector of investments $X = (x_k)$, total investment cost is incurred ex-ante, before the realization of the state of nature, and it equals

$$\sum_k I_k(x_k).$$

After the realization of the state of nature s, total available capacity in state s is

$$K(X, s) \equiv \sum_k x_k K_k(s).$$

We also define the aggregate cost function in state s, given the investments X:

$$C(Q, X, s) = \min\{\sum_k x_k c_k(q_k, s) : 0 \leq q_k \leq K_k(s), \sum_k x_k q_k = Q\}.$$

Moreover, we shall denote by C_Q the marginal production cost. Finally, we shall also make use of the following profit functions:

$$\pi_k(p, s) = \max\{pq_k - c_k(q_k, s) : 0 \leq q_k \leq K_k(s)\}.$$

Example: intermittency. There are two types of production units, green and brown. A brown unit ($k = b$) costs $d_b > 0$ and produces electricity at a constant marginal cost $c_b \geq 0$, up to a fixed capacity K_b. A green unit ($k = g$) produces electricity at a lower marginal cost, that we normalize to zero, up to a capacity $K_g(s)$ which is allowed to depend on s, so as to model intermittency. We shall interpret the state of nature s as a multiplicative weather shock that takes values in $[0, 1]$, with a c.d.f. F. Moreover, possible locations for these units are usually scarce, and we model this by assuming that the investment cost $I_g(x_g)$ is strictly convex in the number x_g of green units. Therefore, with slight abuses of notation we have, in each state s:

$$K_g(s) = sK_g \qquad K_b(s) = K_b \qquad I_b(x_b) = d_b x_b \qquad K(X, s) = x_g s K_g + x_b K_b.$$

Moreover, the efficient manner to satisfy a demand e is to use the green units first, and then the brown units; so that, if $Q \leq K(X, s)$, we have

$$C(Q, X, s) = c_b \max(Q - s x_g K_g, 0).$$

Demand: There is a mass of (possibly heterogenous) consumers, aggregated into a representative consumer, who derives a gross surplus $v(e, s)$ from the consumption of a quantity e of electricity in state s.

Assumption 1 *The gross surplus function v is strictly concave in e, for each realization of the exogenous state of nature s.*

The consumer gets electricity thanks to a retail contract. For technological reasons, it is impossible to partially disconnect a consumer: either a consumer sees his demand satisfied, or he is disconnected and gets zero. Moreover, we assume it is too costly for the consumer to react in real time to new information about prices; while he may adapt his consumption in real time to different states of nature, such as changes in temperature. These elements create a strong rigidity in the management of demand. In particular, retail contracts simply specify a retail price $\overline{p}$, set before the realization of the state of nature s, and a fixed fee A. After s is realized, the consumer observes its value and chooses his demand $D(\overline{p}, s)$ from the maximization of the net surplus[4]

$$v(e, s) - \overline{p}e.$$

Thanks to the concavity of v, this demand function is decreasing with the retail price $\overline{p}$. In what follows, D_p will denote the derivative of the demand with respect to price. We shall also make use of the elasticity of demand $\varepsilon = -\frac{pD_p}{D}$.

Finally, when a consumer is disconnected, he obtains a gross surplus $v(0, s)$, and does not pay anything; so that overall his expected net surplus is

$$m(v(e, s) - \overline{p}e) + (1 - m)v(0, s).$$

Now, suppose that the global consumption e is reduced by one unit, so that one has to allocate this reduction across consumers. We assume that the only way is to disconnect some consumers, and that this disconnection is uniform across types of consumers. Hence, the regulator disconnects the same proportion y of each type, so that $y = 1/e$, and when disconnected the representative consumer experiences a loss $v(e, s) - v(0, s)$. The associated aggregate social loss is called the Value of Lost Load, and it equals

$$\ell(e, s) = \frac{v(e, s) - v(0, s)}{e}.$$

Notice that ℓ is decreasing with e, by concavity of the surplus function v. It represents the social loss associated to a reduction by one unit of electricity supply, when all consumers are connected and face the same retail price,[5] and when our assumption of rigid demand management holds.[6] We shall assume that ℓ is high, and lies above any conceivable marginal cost of production.

[4]At this stage, it is interesting to note that the use of such contracts creates an adverse selection problem for the retailers, since a consumer may consume more or less during peak hours when the wholesale price of electricity is high. We shall ignore this problem in this article, either because the consumer's type is public, or beccause all consumers are identical, and we refer to Joskow and Tirole (2006) for a study of these effects.

[5]If only a proportion $m < 1$ of consumers are connected to the grid, then this social value must be multiplied by $1/m$.

[6]Notice that, if demand could be flexibly managed, then the VOLL would be $v'(e)$, which is lower than $\ell(e)$. Alternatively, one could possibly choose which consumers to disconnect: then one would disconnect those with the lowest individual value of lost load, creating a social loss that lies below $\ell(e)$. The point here is that the definition of the VOLL depends on assumptions on the management of demand. The notion of VOLL is thus more complex than it may seem at first glance. As discussed in Joskow and Tirole (2007), it also depends for example on whether consumers may undertake precautionary measures. Our definition also assumes that all consumers optimize their consumption and face the same unit price.

Balancing demand and supply: Demand and supply of electricity must be equalized at any time, and this may raise some difficulties, in particular when demand exceeds the available capacity. In such a case, the rigidity of demand imposes that some rationing takes place, which means that some consumers must be cut off from the grid. But disconnecting some consumers sometimes implies to disconnect simultaneously some of the production units. This also re-routs electricity towards transmission lines that may already be congested, thereby triggering additional disconnections. One ends up with a rolling blackout with rationing, or even a systemic blackout during which entire parts of the grid are out of function during several hours, or even days.

For the sake of simplicity, we do not model this process explicitly. We simply assume that once s is realized, the process ends up with a share m of consumers whose demand e is satisfied, thanks to the production from a share n of production units; the other consumers and production units are disconnected and thus idle. We make the following three assumptions.

We assume that disconnection applies uniformly, both for consumers and for production units, so that disconnection affects only the size of each side of the market, and not its composition. We can thus reason as if there were a mass n of identical producers, each endowed with the cost function C defined as above; and a mass m of identical consumers, each endowed with the surplus function v defined above. This has two consequences.

First, each producer has to satisfy the demand of $r = m/n$ consumers, and therefore the realized total cost takes a simple form, since it equals $nC(re)$.

Second, feasibility implies the following constraints:

$$e = D(\overline{p}, s) \qquad K = K(X, s) \qquad me \leq nK.$$

Our second important assumption aims at neutralizing scale effects: we assume that the final share n of connected producers only depend on the state of nature s, and on the following **capacity index**, equal to the ratio between total available capacity and total demand:

$$\kappa = \frac{K}{e}.$$

Our last important assumption is commonly used in the literature, and allows to isolate natural properties for the functions m, n, and r. When the capacity index is above 1, it is reasonable to assume that everything goes well: all producers are connected, and all consumers get electricity. Indeed, recall that exogenous incidents that affect only producers are already taken into account by the dependence of capacities on the state of nature, so that this assumption amounts to ignoring exogenous incidents that affect some consumers, such as disconnections due to a storm, but that are not related to an imbalance on the market.

More interesting is the case when the capacity index is below 1. Then some consumers must be disconnected, so as to fulfill the inequality $me \leq nK$. But as we explained above, such a process sometimes requires to also disconnect some producers, or may even trigger accidents on the grid, and this means that both m and n can take values below 1. What we assume here is that while the share n of connected producers is exogenously given, the share of connected consumers is chosen by the regulator to be the maximum value consistent with the above inequality.[7] Hence, we obtain $me = nK$, or equivalently $r = \kappa$.

To summarize, we have, if $\kappa > 1$:

$$n(\kappa, s) = m(\kappa, s) = r(\kappa, s) = 1,$$

while if $\kappa \leq 1$:

$$r(\kappa, s) = \kappa \qquad m(\kappa, s) = \kappa n(\kappa, s).$$

The only exogenous function is thus n, while m and r are consequences of the regulator's behavior. Finally, we say that there is a blackout when $\kappa \leq 1$, that the blackout is rolling when $n = 1$, and that it is systemic when $n < 1$.

2 Optima

Let us define optima within this general model, keeping in mind that additional assumptions will be needed later on to ensure the existence and uniqueness of these allocations. An allocation specifies investments X and consumptions $e(s)$ in each state. An optimal allocation is chosen to maximize total expected surplus. The constraint of rigid demand management requires the existence of a retail price $\overline{p}$ such that in each state s we have

$$e(s) = D(\overline{p}, s), \tag{1}$$

[7] While this assumption seems reasonable, it is worthwhile to note that the regulator could as well decide to disconnect more consumers than is needed. We leave this possibility for future research.

a constraint to which we associate the multiplier $\alpha(s)$: this multiplier thus measures the impact of the rigid demand management constraint in state s. We will see later that it may take positive or negative values. Moreover, by definition of the capacity ratio we have in each state s:

$$\kappa(s)e(s) = K(X, s), \tag{2}$$

a constraint to which we associate the multiplier $\beta(s)$: this multiplier thus measures the value of one additional unit of capacity in state s, and we expect it to be positive. The expected consumer surplus is

$$E[m(\kappa(s), s)v(e(s), s) + (1 - m(\kappa(s), s))v(0, s)],$$

from which we have to subtract the expected total cost

$$E[n(\kappa(s), s)C(r(\kappa(s), s)e(s), X, s)] + \sum_k I_k(x_k).$$

From now on, we often omit the argument s. Recall also that we have $m = rn$. Therefore, total surplus, bar an additive constant, simplifies to:

$$E[n(\kappa)\,(r(\kappa)(v(e) - v(0)) - C(r(\kappa)e))] - \sum_k I_k(x_k),$$

and we have to maximize it with respect to e, κ, and p, under the constraints (1) and (2). Remember also that we have $r = \min(\kappa, 1)$.

We now derive the necessary first-order conditions, first with respect to κ in each state: dividing by e, and omitting most arguments, we obtain

$$\frac{\partial n}{\partial \kappa} r[\ell(e) - AC(re)] + n\frac{\partial r}{\partial \kappa}[\ell(e) - C_Q(re)] = \beta. \tag{3}$$

In words, an increase in κ changes the proportion $m = nr$ of connected consumers, either through a change in n, or through a change in r. A change in the size n only changes the scale of production, and thus involves the difference between the value of an additional unit ℓ, and its average cost AC. By contrast, a change in the production re from each producer involves the difference between ℓ and the marginal cost. The first term only appears when an increase in capacity modifies the size of a systemic blackout, while the second term appears when this additional capacity allows to connect more consumers.

Now, when $\kappa > 1$ there is no blackout, both n and r are constant, and therefore $\beta(s) = 0$. And when $\kappa \leq 1$, there is a blackout, and we expect β to be positive if n is increasing with κ. The condition with respect to $e(s)$ in each state gives the value of the other multiplier:

$$nr(\overline{p} - C_Q(re)) - \beta\kappa = \alpha. \tag{4}$$

Indeed, an increase in the electricity consumption impacts the marginal surplus of each of the $m = nr$ connected consumers, who can now consume more; but it requires to invest in capacity to keep κ constant. Typically, the multiplier α is positive when there is no blackout and costs are low, but may be negative when the rigid demand management constraint forbids a desirable reduction of electricity consumption per capita.

The condition with respect to the retail price is simply

$$E[\alpha D_p] = 0. \tag{5}$$

It will be used to determine the optimal retail price, using (4). Finally, observe that the reduction in cost in state s due to an additional unit of type k equals the profit from this unit, at the global marginal cost:

$$\frac{\partial C}{\partial x_k}(Q, X, s) = -\pi_k(C_Q(Q, X, s), s),$$

where x_k stands for the number of units of type k. Therefore, the first-order condition with respect to investment in a unit of type k writes (assuming it is optimal to use at least one such unit):

$$I'_k(x_k) = E[n\pi_k(C_Q(re))] + E[\beta K_k]. \tag{6}$$

The first two terms measure the private incentives to invest, assuming electricity is priced at the global marginal cost (though we shall soon explain this cannot be the case in a competitive equilibrium). The last term is the expected social value of this additional capacity. The first consequence is that in a market equilibrium, pricing electricity at marginal cost is unlikely to exert sufficient incentives for investment in capacity – unless blackouts never occur. Notice that also the social value of each unit of capacity depends on the state of nature, as expressed in (3). When capacity is independent from s, it should receive a subsidy of $E\beta$ per unit; when capacity is variable, this subsidy should be corrected to take into account the correlation between capacity and its social value

β. This second consequence suggests that additional capacity should be subsidized only when it is indeed available during a blackout, and has a high social value. This will have some consequences for intermittent sources, as we shall see.

3 Competitive Equilibria (with blackouts and price cap)

We now turn to the notion of competitive equilibria we use in this paper. At this stage, it is important to underline that wholesale electricity markets face a structural difficulty: namely, that it is impossible to balance demand and supply when the available capacity is less than the realized demand. The idea that agents are price-takers becomes awkward when there is no price that would balance the market. Our notion of competitive equilibria thus has to be augmented by regulatory interventions.

The first regulatory reaction to an imbalance on the market is to randomly disconnect some consumers, as we have just explained. But this is not enough: when the minimum number of consumers is disconnected, every producer remains indispensable to the satisfaction of the remaining consumers. For the modeler, a first possibility is to rely as much as possible on the concepts from competitive equilibria. If each producer is infinitesimal, and take prices as given, then in equilibrium electricity can be priced at marginal cost, even when the capacity constraint binds,[8] and even when there is a blackout.

Unfortunately, this intellectual construction does not seem to fit the reality of electricity markets: prices tend to increase beyond control before enough consumers are disconnected. The second typical intervention consists in imposing an administrative price cap P, used to settle transactions between producers and retailers, and chosen high enough so as to ensure that all available capacity is put to use.

In our model, we introduce these two regulatory interventions as follows. Disconnection is modeled through the use of functions m, n, and $r = m/n = \min(\kappa, 1)$. If $\kappa > 1$, then there is excess capacity, producers act as price-takers, and electricity is priced at marginal cost. If $\kappa \leq 1$, then all non-infinitesimal producers become indispensable, and they can ask for an arbitrary high price in exchange of its production. In such a case, we assume that the electricity price is capped at the administrative value P.

[8] In such a case, marginal cost is computed at the left of the value of available capacity, i.e., $\frac{\partial C}{\partial Q}(K^-)$.

This modeling choice can be seen as a simple manner to model imperfect competition. Regrettably, it introduces a discontinuity between situations when $\kappa \leq 1$, and situations when $\kappa > 1$. Notice however that this discontinuity would still appear if one were to model more precisely imperfect competition between producers. In such an hypothetical case, there would appear a threshold for κ such that, when κ exceeds this threshold, then some strategic producers would withdraw capacities so as to increase the price up to the cap P; while if κ is just below the threshold, pricing would obey the standard rules of, say, Cournot pricing. Whether such opportunistic behaviors are likely to be discouraged by the fear of public reactions is difficult to assess; the point here is that this discontinuity is a natural consequence of the possibility of blackouts, and of strategic behavior on the supply side.

This being clarified, let us turn to the derivation of competitive equilibria, for a given price cap P. As before, one has to find an allocation (κ, e) in each state of nature, together with endogenous equilibrium prices which are the retail price $\overline{p}$ and the wholesale prices $p(s)$; only the latter is allowed to depend on the state of nature. The retail price determines demand, through the equality (1). Constraint (2) still holds, by definition of the variables involved. Wholesale prices are determined in each state, according to two possible rules. First, if $me < nK$, or equivalently $\kappa > 1$, then the market efficiently balances demand and supply, and we have

$$p(s) = C_Q(e(s), X, s). \tag{7}$$

Otherwise, we have $\kappa \leq 1$, and all connected production units are indispensable. Then the administrative price cap P prevails, and we assume it is high enough to ensure that these units produce up to capacity, i.e.,

$$p(s) = P > C_Q(K(X, s), X, s). \tag{8}$$

At this stage, it is important to mention that the price cap P will in general not suffice to fully decentralize optima. Moreover, it is often argued that this cap should be further reduced, for two reasons that are not explicitly considered in our model: first, because retailers want to reduce the risk they bear by buying at the risky wholesale price to sell at the fixed retail price; second, because with a lower price cap producers benefit less from their market power (see Wolak, 2013, and Fabra, 2018, among many others). We shall thus often take P as exogenously given, at a low level compared to the reference level which is the Value of Lost Load.

This creates two additional distortions, requiring two additional instruments:

• The well-known "missing money" issue underlines that with a low cap, producers are not rewarded sufficiently for their investments. One should thus introduce additional incentives for capacity creation. We thus introduce a subsidy σ_k for each unit of type k.

• A low cap also reduces the price retailers have to pay for getting electricity, thereby reducing the competitive retail price. One has to correct this distortion by creating a tax τ on retail electricity.

Overall, the regulation of electricity markets appears quite complex. Because retail contracts are so specific, one has to allow for disconnections, and to introduce a price cap P, a retail tax τ, and capacity subsidies σ_k. We now examine how to determine the optimal values for these instruments.

Retailers buy electricity at the wholesale price, to resell it to consumers at the retail price, including the retail tax. Being competitive intermediaries with constant returns to scale, they take wholesale prices as given, and they get zero-profit on average. However, one has to be careful: the contract which is sold to a consumer specifies that he can buy whatever quantity he wishes, in every state of nature, at the price $\overline{p}$. This contract itself can be sold, for a fixed fee A. Competition should then select the contract $(A, \overline{p})$ that maximizes the consumer's payoff

$$E[m(\kappa(s), s)(v(D(\overline{p}, s)) - \overline{p}D(\overline{p}, s))] - A,$$

under the constraint that profits be nonnegative:

$$A + (\overline{p} - \tau)E[m(\kappa(s), s)D(\overline{p}, s)] \geq E[p(s)m(\kappa(s), s)D(\overline{p}, s)].$$

By making this constraint bind, one obtains the classical result that the use of an access charge promotes efficiency, since the equilibrium contract gives zero-profit to retailers and maximizes over $\overline{p}$ the surplus

$$E\left[m(\kappa(s), s)[v(D(\overline{p}, s), s) - (p(s) + \tau)D(\overline{p}, s)]\right].$$

This result appears also in Joskow and Tirole (2006, 2007). These articles also underline that these retail contracts raise difficulties of their own when consumers are heterogenous, since different consumers with different load profiles correspond to different expected costs for the retailer. If the type of consumers is observed, then the efficiency result extends; but one now faces several retail contracts. For the sake of simplicity, we abstract from these difficulties, and we only consider one retail contract, for which $\overline{p}$ maximizes the expected aggregate surplus, with the following necessary and sufficient first-order condition:

$$E[m(\kappa(s), s)(\overline{p} - p(s) - \tau)D_p(\overline{p}, s)] = 0. \tag{9}$$

and the access charge ensures that retailers' profits are zero, for each customer.

Finally, every investment should balance the marginal cost of an additional unit with the additional expected revenues, so that, once more assuming an interior solution, and taking the capacity subsidy into account:

$$\text{for all } k \qquad I_k'(x_k) = E[n(\kappa(s), s)\pi_k(p(s), s)] + \sigma_k. \tag{10}$$

4 Decentralization

The next step consists in comparing equilibrium allocations to optima, and, if these allocations differ, to determine the regulatory instruments that are needed to decentralize optima. To do so, recall that in both cases one must have

$$\kappa(s)e(s) = K(X, s) \qquad e(s) = D(\overline{p}, s).$$

There remains to compare (3)-(4)-(5)-(6) to (7)-(8)-(9)-(10). To do so, we examine different cases in turn. As already discussed, competition on the retail side is assumed efficient, even under a rigid management of demand; the key difficulty will lie in the fact that blackouts threaten the efficiency of investments.

4.1 No blackout

Let us begin by the simplest case, in which it is assumed that blackouts cannot happen: so that one must have $m = n = 1$, and consequently $r = 1$. Note that for such an outcome to be feasible, it must be that κ is above one with certainty, or equivalently:

$$\text{For all } s \qquad K(X,s) > D(\overline{p},s). \tag{11}$$

Keeping this condition in mind, our study is easily performed. For optima, the pair of conditions (3)-(4) reduces to

$$\beta(s) = 0 \qquad \alpha(s) = \overline{p} - C_Q(D(\overline{p}))$$

so that the remaining conditions (5)-(6) are:

$$\overline{p}ED_p = E[C_Q(D)D_p]$$

$$\text{For all } k \qquad I'_k(x_k) = E[\pi_k(C_Q(D))].$$

Turning now to equilibria, (7) always applies, while (8) is irrelevant. And (9) and (10) turn out to be identical to the above conditions, setting to zero the retail tax and the capacity subsidy. We conclude that in the absence of blackouts, competition is efficient, and regulation is not needed.

Let us acknowledge that this result stems directly from our assumption that producers behave in a competitive way, i.e., they take prices as given, when there is excess capacity. We thus depart from an important strand of literature by ignoring market power. This may be justified if this excess capacity is high enough.

But the main difficulty here lies with the fact that condition (11) requires that this excess capacity is left idle with probability one. This might indeed constitute the best thing to do, for example if demand and available capacity are not too volatile. However, creating a significant excess capacity, together with a commitment to use it only to punish collusive behaviors, does not seem to form a credible policy; and this excess capacity would need to be quite high if the difference between demand and available capacity is highly volatile. We conclude that blackouts are unavoidable.

4.2 Only rolling blackouts

We now investigate the case when producers are never disconnected during a blackout: $n = 1$. Then either $\kappa > 1$, and then demand is satisfied, and electricity is priced at marginal cost; or $\kappa \leq 1$, and demand is rationed by a factor κ, and the full capacity

$K(X,s)$ is used, and electricity is priced at the price cap P. We therefore have $m = r = \min(\kappa, 1)$.

For optima, the pair of conditions (3)-(4) now take different forms according to whether there is a blackout or not: if $\kappa > 1$, we obtain

$$\beta(s) = 0 \qquad \alpha(s) = \overline{p} - C_Q(D(\overline{p}))$$

while if $\kappa \leq 1$ we get

$$\beta(s) = \ell(D(\overline{p})) - C_Q(\kappa D(\overline{p})) \qquad \alpha(s) = \kappa(\overline{p} - \ell(D(\overline{p}))).$$

The first optimality condition (5) now becomes

$$E[1_{\kappa>1}(\overline{p} - C_Q(D))D_p + 1_{\kappa\leq 1}\kappa(\overline{p} - \ell(D))D_p] = 0.$$

It should be compared to the equilibrium condition (9):

$$E[1_{\kappa>1}(\overline{p} - \tau - C_Q(D))D_p + 1_{\kappa\leq 1}\kappa(\overline{p} - \tau - P)D_p] = 0.$$

Recall that κ is the ratio of available capacity to demand, so that we can usefully introduce the demand elasticity ε. The two conditions thus coincide if and only if

$$PE[1_{K<D}K\varepsilon] + \tau E[\min(K,D)\varepsilon] = E[1_{K<D}K\varepsilon\ell]. \tag{12}$$

The last condition for optimality is (6), which writes:

$$\text{for all } k \qquad I_k'(x_k) = E\pi_k(C_Q) + E[1_{\kappa\leq 1}(\ell - C_Q)K_k].$$

Recall that when κ is below 1 all available capacity is used. Therefore, this condition can be rewritten as

$$\text{for all } k \qquad I_k'(x_k) = E[1_{\kappa>1}\pi_k(C_Q)] + E[1_{\kappa\leq 1}(\ell K_k - c_k(K_k))].$$

This condition can be compared to the last condition for equilibrium (10):

$$\text{for all } k \qquad I_k'(x_k) = E[1_{\kappa>1}\pi_k(C_Q)] + E[1_{\kappa\leq 1}\pi_k(P)] + \sigma_k.$$

Now, the comparison of these two conditions is easily performed: they coincide if and only if

$$\text{for all } k \qquad E[1_{\kappa\leq 1}\pi_k(P)] + \sigma_k = E[1_{\kappa\leq 1}(\ell K_k - c_k(K_k))].$$

Recall that capacity is fully used when the price cap applies, so that this equality becomes

$$\text{for all } k \qquad PE[1_{K<D}K_k(s)] + \sigma_k = E[1_{K<D}\ell K_k(s)]. \tag{13}$$

Conditions (12) and (13) ensure that the conditions for optimality coincide with the conditions for equilibria. The first positive result is the following:

Claim 1 *Suppose that systemic blackouts are excluded. For any price cap P, it is possible to find a retail tax and capacity subsidies that decentralize an optimal allocation.*

Two qualifiers here: first, we have assumed from the start that the price cap is high enough to incentivize producers to produce up to capacity in case of blackouts. Second, the conclusion relies only on the comparison of first-order conditions, while the uniqueness of equilibria and of optima cannot be taken for granted.

The second positive result relates the price cap to the value of lost load, in the absence of intermittency, and when the elasticity of demand is a constant:

Claim 2 *Suppose that systemic blackouts are excluded. Suppose also that the elasticity of demand is a constant, and that there is no intermittency, so that capacities do not depend on the state of nature. Then one can implement an optimum with a single instrument, by setting the price cap equal to the expectation of the VOLL, conditional on a blackout:*

$$P = E[\ell(D(\overline{p}, s), s)|K < D(\overline{p}, s)].$$

We are back to standard reasonings: the value of lost load should determine the value of the price cap. Note however that the VOLL is computed conditional on a blackout here, and in the absence of intermittency blackouts happen when demand is high, so that the VOLL is also high.

4.3 General case: systemic and rolling blackouts

In the general case, for some states of the world it happens that $n < 1$ in which case a systemic blackout occurs. In this general setting, either $\kappa > 1$, and then demand is satisfied, and electricity is priced at marginal cost; or $\kappa \leq 1$, and demand is rationed by a factor κ. When $n = 1$, the full capacity $K(X, s)$ is used, and $m = r = \kappa$; whereas when

$n < 1$, the available capacity $nK(X,s)$ is used, and $m = nr = n\kappa$. In case of a blackout (rolling or systemic), electricity is priced at the price cap P. As a consequence, in this general case, $r = \min(\kappa, 1)$ and $m = \min(n\kappa, \kappa, 1)$.

For optima, the pair of conditions (3)-(4) now take different forms according to whether there is a blackout or not: if $\kappa > 1$, we obtain

$$\beta(s) = 0 \qquad \alpha(s) = \overline{p} - C_Q(D(\overline{p}))$$

if $\kappa \leq 1$ and $n = 1$ we get

$$\beta(s) = \ell(D(\overline{p})) - C_Q(\kappa D(\overline{p})) \qquad \alpha(s) = \kappa(\overline{p} - \ell(D(\overline{p})))$$

last, if $\kappa \leq 1$ and $n < 1$ we get

$$\begin{aligned} \beta(s) &= n\left(\ell(D(\overline{p})) - C_Q(\kappa D(\overline{p}))\right) + \frac{\partial n}{\partial \kappa}\kappa\left(\ell(D(\overline{p})) - AC(\kappa D(\overline{p}))\right) \\ \alpha(s) &= n\kappa\left(\overline{p} - \ell(D(\overline{p}))\right) - \frac{\partial n}{\partial \kappa}\kappa^2\left(\ell(D(\overline{p})) - AC(\kappa D(\overline{p}))\right). \end{aligned}$$

The optimality as well as the equilibrium conditions should be computed and compared.

Bibliography

S. Ambec and C. Crampes (2019), Decarbonizing Electricity Generation with Intermittent Sources of Energy, *Journal of the Association of Environmental and Resource Economists,* vol. 6, No. 6, pp. 1105-1134.

S. Auray, V. Caponi, and B. Ravel (2019), Price Elasticity of Electricity Demand in France, *Economie et Statistique-Economics and Statistics,* 513, 91-103.

P. Cramton, A. Ockenfels, and S. Stoftc (2013), Capacity Market Fundamentals, *Economics of Energy and Environmental Policy,* vol. 2, No. 2, pp. 27-46.

N. Fabra (2018), A primer on capacity mechanisms, *Energy Economics,* vol. 75, pp. 323-335.

R. Gerlagh, M. Liski, and I. Vehviäinen (2022), A price control mechanism for a persistent supply shock, working paper.

P. Joskow and J. Tirole (2006), Retail electricity competition, *RAND Journal of Economics,* Winter, Vol. 37, No. 4, pp. 799-815.

P. Joskow and J. Tirole (2007), Reliability and competitive electricity markets, *RAND Journal of Economics,* Vol. 38, No. 1, Spring, pp. 60-84.

P. Joskow (2021), From hierarchies to markets and partially back again in electricity: responding to decarbonization and security of supply goals, *Journal of Institutional Economics,* pp. 117.

F. A. Wolak (2013), Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables, *Review of Economics and Institutions,*, Winter, Vol. 4, No. 1, pp. 1-42.

F. A. Wolak (2021), Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables, working paper.