WORKING PAPERS

N° 1555

August 2025

“Optimal Regulation of Electricity Provision with Rolling and Systemic Blackouts ”

Catherine Bobtcheff, Philippe De Donder and François Salanié

Toulouse School of Economics

Optimal Regulation of Electricity Provision with Rolling and Systemic Blackouts*

Catherine Bobtcheff† Philippe De Donder‡ François Salanié§

July 11, 2025

Abstract

We set up a static model of electricity provision in which delivery to consumers is only imperfectly reliable. Blackouts can be either rolling or systemic; in both cases a price cap becomes active on the wholesale market. We show that for any given value of the price cap, one can decentralize optimal allocations thanks to two types of regulatory instruments: a retail tax, and capacity subsidies. Some properties follow. If demand is affected by multiplicative shocks only, capacity subsidies are exactly financed by the revenues from the retail tax. If moreover the distribution of systemic blackouts is exogenous, a price cap is sufficient, provided it is set at the value of lost load. In all other cases, all instruments are needed, and capacity subsidies need to be differentiated, based on the correlation between available capacity and its social value. We also discuss the impacts of a carbon tax on supply, demand, and optimal regulation.

Keywords: electricity, reliability, renewables, climate change.

JEL Classification: D24, Q41, Q42, Q48.

*We thank Stefan Ambec, Claude Crampes, Shefali Khanna, participants to the 13$^{\text{th}}$ Toulouse Conference on the Economics of Energy and Climate, the 27$^{\text{th}}$ Annual Conference of the European Association of Environmental and Resource Economists (EAERE, Rimini), the X$^{\text{th}}$ Conference of the Spanish-Portuguese Association of Natural and Environmental Resources Economics (AERNA, Barcelona), 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. The authors gratefully acknowledge funding from the French National Research Agency (ANR) under its Programme d'Investissement d'Avenir ANR-17-EURE-0001 and ANR-17-EURE-0010.

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

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

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

1 Introduction

Climate change, and the policies undertaken to mitigate it, are deeply transforming the energy sector. The drive away from fossil fuels has led to the development of so-called intermittent sources that rely on sunlight or wind to produce electricity. Other energy sources 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 transportation to heating to cooling, and also for many industrial processes.

This electrification of uses 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 in continuous time. This balance is difficult to manage when a significant part of supply is intermittent; when demand depends also on weather, and cannot be managed in real time; and when additional pressure is put on production units and on the grid because of more extreme climatic events. Reliability, including the possibility of major blackouts, is thus likely to become a major concern in the future, both for households and for industrial or commercial processes. In the absence of alternative energy sources, users will face a risk of disconnection. The February 2021 crisis in Texas, and the April 2025 Spanish blackout, illustrate perfectly the costs and hazards of the unreliability of electricity production and distribution.

This paper aims at analyzing these risks, and how they impact the optimal regulation of electricity markets. Our model of the electricity sector is inspired by Joskow and Tirole (2007). It is a static model that allows for heterogenous producers, incorporates intermittency and arbitrary shocks to costs, capacities and demand. We add an original modeling of disconnections, and of rolling and systemic blackouts. Our methodology is straightforward: We characterize optimal allocations (i.e., investments, and productions and consumptions in each state of nature), and we show how to decentralize these optima thanks to a regulatory intervention; in particular, we derive explicit formulas for the optimal values of regulatory instruments. We conclude the paper with an application to carbon taxation.

Let us begin by summarizing the main steps in our analysis. We emphasize that difficulties with the electricity market mainly originate on the demand side. At the most local level, agents consume electricity by connecting devices to their own private loop, which is itself connected to the grid. By switching on and off these devices, agents determine their electricity demand. For technical reasons, each individual demand is either fully satisfied, or not at all: partial rationing of a consumer's demand is so far impossible.[1] Another difficulty is behavioral: because attention is costly, the typical consumer is not able to react in real time to changes in the price he pays.[2]

These features severely constrain retail contracts. Strikingly, retailers all over the world have ended up selling the same type of contract: in exchange for a fixed fee, this contract allows consumers to demand the quantity they wish at a pre-specified retail price. This rigidity in the management of demand makes it impossible to balance the wholesale electricity market when demand exceeds available capacity. In such a case, each producer becomes indispensable, making the surge in prices irresistible, while doing little to reduce the imbalance. Then the only way to balance the market is to ration demand through disconnection of consumers. Interestingly, regulators in charge of the management of these blackouts typically choose to disconnect the minimum number of consumers consistent with the equality between demand and available capacity. By doing so, they perpetuate the market power of producers since all of them remain indispensable; we stick to this case in our model.[3] Moreover, in order to settle transactions between producers and retailers at a reasonable price, regulators typically impose a cap on the wholesale price. The value for this cap varies from country to country.[4]

[1] This would require setting priorities at the level of each device, or installing 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] This is not the case for some industrial users. We ignore this possibility in this article, for the sake of simplicity.

[3] Regulators display a willingness to avoid orderly blackouts which seems out of proportion with standard estimates of the cost of these episodes; on this point, see Wolak (2013) and Joskow (2022, pp. 5-6).

[4] Since 2022, the cap is set at €3,000 per MWh in Europe. The price cap in Texas was particularly high in 2021, at $9,000; it became binding in February 2021. This value was considerably higher than the average yearly price of $20 in 2020; still, the market could only be balanced thanks to demand rationing.

Unfortunately, setting a price cap creates two distortions, that can be corrected thanks to two new regulatory instruments. On the retail side, retailers benefit from a capped wholesale price of electricity, and therefore the competitive retail price is reduced: this may justify imposing a tax on retail electricity. On the supply side, the producers' revenues 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; we deal with it by relying on subsidies to capacity.

In elementary models, the missing money problem can be fully solved by setting the price cap at the level of the Value of Lost Load (VOLL), defined as the social cost of a power outage: indeed, such a level provides the right incentives to investors and producers. However, this intuitive argument raises a number of difficulties. The definition of the VOLL is not precise enough, as its value may for example depend on whether the outage is anticipated or not,[5] and its measurement raises numerous difficulties.[6] In our model, it is also dependent on contingencies such as temperature: a freezing cold makes heating indispensable to consumers. Finally, there are good reasons for reducing the price cap below the VOLL. With a high price cap, retailers are put into an uncomfortable situation: they sell electricity at a fixed retail price, while they buy it at a very volatile wholesale price. Even if risk-neutrality is assumed, when wholesale prices rise retailers may end up in a situation in which bankruptcy is the only option, as observed recently in the UK when gas prices surged.[7] Setting a lower price cap is thus a manner to reduce the risk borne by retailers, and ultimately by consumers when these retailers go bankrupt. Moreover, it has long been argued that lowering the price cap reduces the exercise of market power by dominant producers.[8] While these considerations are left unmodeled in this paper, they motivate our choice of taking the price cap as given, and deriving the associated optimal regulation.[9]

[5]See Joskow and Tirole (2007) for a precise discussion.

[6]These issues are analyzed in Gorman (2022).

[7]See Haar (2021) for a discussion of retailers' difficulties in the UK.

[8]See Wolak (2013) and Fabra (2018), among many others.

[9]One might envision more complex interventions, such as making the price cap or the level of rationing contingent on market information ex-post, instead of being set ex-ante. Gerlagh, Liski and Vehviläinen (2024) allow for such possibilities in a stylized model of demand rationing with responsive and non-responsive consumers.

To summarize, our view is that the inability to flexibly manage demand in real time creates a market failure: it is sometimes impossible to balance the market. This in turn implies a variety of regulatory interventions: one has to set a price cap, and to ration demand through blackouts. The price cap in turn requires to tax the retail price, and to subsidize capacity. We distinguish two types of blackouts. In a rolling (or orderly) blackout, all available capacity is allocated uniformly to consumers, who then face the same probability of not being served. In a systemic blackout, the allocation of the available capacity to consumers puts some additional stress on the transmission grid, triggering disconnections from the grid of some production units; such episodes can have dramatic consequences.

The point is that the types and sizes of blackouts are not only determined by exogenous shocks. They also depend on the existence of excess capacities that would allow to hedge against unanticipated surges in demand, or breakdowns in production, or incidents in the grid. We model this by allowing the impact of a blackout to depend on what we call the capacity index, namely the ratio of available capacity to aggregate demand. This is a plausible measure of the tension on the grid, which moreover neutralizes scale effects. Hence, every new investment in capacity, and every reduction in demand, contributes to higher values of the capacity index, and this in turn reduces both the likelihood and the size of the two types of blackouts. But this improvement in the performance of the grid is a public good likely to be under-provided in the absence of regulation. To remedy this inefficiency, one is once more led to use our two instruments, i.e. subsidies to capacity and retail tax.

Our first contribution is thus to set up a general model of the electricity sector, with heterogenous producers, arbitrary shocks on both supply and demand, and different types of blackouts. Efficiency is defined as the maximization of total surplus. Efficient investments, productions, and consumptions are characterized through simple first-order conditions. We propose an extended notion of competitive equilibria to study the decentralization of efficient allocations: firms are price-takers, unless demand is rationed, in which case they become indispensable and behave strategically. We finally derive the optimal regulation, for arbitrary values of the price cap.

We now state our results. The main one characterizes the optimal regulation: we show that for any value of the exogenous price cap one can decentralize the optimal allocations using only two types of instruments: a retail tax, and differentiated subsidies for capacity. Our model thus provides a general normative analysis for the electricity sector, which to the best of our knowledge has been absent so far.

The model also emphasizes the symmetry between demand and supply: reducing demand is as important as fostering investments. This can be seen for the missing money problem: the price cap reduces the payments to producers, but it also reduces the costs borne by retailers when buying electricity on the wholesale market, and this distortion should also be corrected. Similarly, as soon as the size and occurrence of blackouts increase with demand and decrease with installed capacity, then capacity should be subsidized, and demand should be taxed. These simple arguments do not appear to be present in the literature, from Joskow and Tirole (2007) to Léautier (2016) to Elliott (2024), nor in Borenstein, Bushnell and Mansur (2023)'s survey of reliability.[10]

The optimal retail tax includes two expected terms, one for the missing money problem and one for the public good problem. Both terms are decreasing with the price cap: indeed, the cap and the retail tax are substitutes since both lead to a reduction in demand. It is then possible to pinpoint a value for the price cap such that the retail tax is nil; but we show that this value of the price cap exceeds the VOLL, in contradiction with the actual practice of regulatory bodies. This justifies the use of a retail tax, even in the absence of environmental issues such as global warming, and even in the absence of blackouts.

The optimal subsidies to capacity share a similar additive structure. For production plants with a constant capacity, the subsidy simply applies to each unit of nominal capacity, independently of the technology. But for intermittent sources, the value of the subsidy depends not only on the nominal capacity, but also on how the effective capacity varies across states of nature. Therefore, the subsidy to solar is reduced if solar panels produce electricity only when its social value is low. The formulas we provide make it possible to evaluate this effect with precision.

[10]An exception is Borenstein and Holland (2005)'s discussion of the choice between a retail tax and capacity payments. But their model has no intermittency and no blackouts, and these assumptions imply that only one regulatory instrument is needed for efficiency.

Formally, the optimal values of these instruments depend not only on expected values, but also on covariances between the elasticity of demand and total capacity, among others. These are complicated objects, and at this point we simplify the problem by assuming that demand is only subject to multiplicative shocks.[11] A striking property follows: at the optimum, the regulator's budget is balanced, i.e., the revenues from the retail tax exactly balance the subsidies to capacity. To the best of our knowledge, this result is new in the literature.[12] It may be valuable as a guideline for public agencies.

We also ask under which conditions setting a price cap equal to the VOLL is a sufficient instrument for optimality, as claimed by the so-called 'energy-only' paradigm. Consider the missing money problem. Suppose that demand is affected by multiplicative shocks only. Then we show that the VOLL is well-defined and does not depend on the state of nature. In such a case, setting a price cap equal to the VOLL indeed solves the missing money problem. However, under more general assumptions on demand, all instruments have a role to play. For example, one can obtain a zero retail tax by setting the price cap at a high value[13]; but capacity subsidies must still be used, because of the correlations between the available capacities and the VOLL. In fact, the only case in which a well-set price cap is sufficient for optimality is when demand is only affected by multiplicative shocks, and there is no intermittency, and blackouts are purely exogenous. These conditions thus severely limit the relevance of the energy-only paradigm.

We conclude the paper by a discussion of carbon taxation. We emphasize that a tax on greenhouse gases emissions impacts both supply (because producers substitute nuclear or renewables for oil, gas, and coal) and demand (because of the electrification of uses: consumers switch to heat pumps or electric cars). We extend our setting to take simultaneously into account both effects. We show through simple simulations that more intermittency on the supply side increases the volatility of available supply, possibly so

[11]This powerful assumption is common in the literature; see for example Joskow and Tirole (2007, property (v) pp. 62), Elliott (2024, pp. 16), or Gowrisankaran, Reynolds and Samano (2016, pp. 1198). This last paper computes the value of intermittent electricity, using US data. It proceeds by solving directly the system operator's problem of maximizing total surplus. More intermittency creates significant costs because storage is not available.

[12]The 2003 working paper version of Borenstein and Holland (2005) offers a discussion on budget-balanced regulations.

[13]Precisely, at the expectation of the VOLL, conditional on demand being rationed, with some weights that take total available capacity into account.

much that the only manner to limit the probability of blackouts is to restrict demand drastically.

We now turn to the related literature. The importance of a flexible management of demand through real-time pricing is underlined in Borenstein and Holland (2005) and Joskow and Tirole (2007). The absence of real-time pricing is a case of market incompleteness (Ferrasse, Neerunjun and Stahn (2022)). It makes demand rigid, and is responsible for the occurrence of blackouts. Our paper thus emphasizes the connection of the management of demand to the issue of reliability, thanks to an original and compact modeling of both rolling and systemic blackouts. As in Keppler (2017), we base our analysis on the identification of such market failures: here, the public good problem of grid reliability, and the fact that some markets are missing.

A key role is played by the function linking the occurrence and size of blackouts to realized demand and available capacity, plus stochastic shocks. This function is in principle identified from existing data, though the econometrician would face the difficulty of establishing the probability of very rare events such as a large, systemic blackouts. Explicit models of blackouts are scarce in the literature. Both rolling and systemic blackouts are studied in Joskow and Tirole (2007), though in a simplified model, and without our optimality result for regulatory instruments. Llobet and Padilla (2018) study investment choices when blackouts foster a reputation loss. Elliott (2024) models only rolling blackouts, with disconnections of the demand side in proportion to the ratio of available capacity over demand. All papers acknowledge the existence of a public good problem associated with reliability, as argued in Abbott (2001).

Being essentially static, the model ignores important determinants of investments that are the focus of much more ambitious works. The fully dynamic model in Elliott (2024) incorporates dynamic investment strategies in a non-stationary environment: this allows to precisely study the timing of investments under various scenarios. It also allows for market power, both at the investment stages and on the wholesale market. A similar emphasis on investment timing can be found in Gowrisankaran, Langer and Zhang (2025). We also ignore the possibility of storage, using for example batteries, as studied in Butters, Dorsey and Gowrisankaran (2024) or Holland, Mansur and Yates (2022).

Another deliberate choice is to reduce the exercise of market power to the case when each producer becomes indispensable, i.e when demand is rationed because no price could possibly balance the market. Our view is that this non-existence of equilibrium price is a characteristic of electricity markets which needs to be studied per se. It is an important source of market power that differs from the day-to-day, Cournot-style market power which admittedly exists on electricity markets, but is in no way different from market power in any other market. Indeed, one may argue market power matters essentially in times of crisis, when the social value of electricity can reach very high levels.[14] Such a stand contrasts with much of the applied literature that aims at recovering information on costs and demand from the observed strategies. Still, it allows us to provide a clear picture of market failures, and of the instruments needed to remedy these failures.[15]

One reason why intermittent sources of energy have been on the increase is that their use does not emit greenhouse gases into the atmosphere. Introducing an optimal tax on emissions in our model would not change any of the results: it would simply modify the unit cost of different technologies. A sub-optimal tax creates more complex distortions, that could be corrected by subsidizing green electricity, and once more by creating a retail tax, as in Ambec and Crampes (2019). The model in Elliott (2024) has three market failures (pollution, inelastic final demand, market power) and two instruments (carbon tax and capacity payment). A carbon tax favors renewables, but makes blackouts more likely; capacity payments exert opposite effects. The model is estimated on Western Australia data in order to fund the right balance between these two instruments. By contrast, we claim that optimality can be reached with a Pigovian carbon tax, and capacity payments and a retail tax set according to our formulas; but we ignore market power issues, as well as the dynamics of investment.

We proceed as follows. The next section presents the model. We then characterize optimal allocations in Section 3, competitive equilibria in Section 4, and a first-best regulation in Section 5. Section 6 presents an analysis of the effects of carbon taxation, as well as some simulations. Section 7 concludes. Proofs and details of the derivations

[14]Interestingly, firms with market power may behave so as to make such episodes more likely. McRae and Wolak (2019) claim to have identified such behaviors on the Colombian electricity market.

[15]Ferrasse, Neerunjun and Stahn (2022) also rely on perfectly competitive equilibria to show how endogenous retail contracts can address the incomplete market problem and eliminate both the non-existence problem and demand rationing. Assuming perfect competition is sometimes illuminating.

are gathered in the Appendix.

2 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 sign retail contracts with consumers. Then, the state of nature s is publicly realized. Ex-post, consumers choose their demand, producers decide what to produce, and the wholesale price is determined on the market, depending on stochastic blackouts and on regulation. Retailers pay the wholesale price to producers and distribute electricity to consumers, or at least to those producers and consumers 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 $c_k^s(q_k)$ in state s, weakly convex in production $q_k \in [0, K_k^s]$. We allow the (finite) production capacity K_k^s to depend on the state s, so as to capture *inter alia* the intermittency due to weather.

Ex-ante, it is possible to invest in an arbitrary number $x_k \geq 0$ of units of type k. To avoid lengthy discussions of corner or integer solutions, we assume that the numbers x_k are strictly positive real numbers. The associated investment cost $I_k(x_k)$ is assumed weakly convex, 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 equals

$$\sum_k I_k(x_k).$$

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

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

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

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

We denote by C_Q^s the marginal production cost, and by AC^s the average cost. Finally, we let π_k be the profit function of a unit of type k:

$$\pi_k^s(p) = \max\{pq - c_k^s(q) : 0 \le q \le K_k^s\},$$

where p denotes the wholesale price of electricity. Also, a useful identity is:

$$\frac{\partial C^s}{\partial x_k}(Q,X) = -\pi_k^s(C_Q^s(Q,X)). \tag{1}$$

Demand. There is a mass of (possibly heterogenous) consumers, aggregated into a representative consumer, who derives in each state s a strictly concave gross surplus $v^s(e)$ from the consumption of a quantity e of electricity. The consumer is aware of the state s, and sets consumption accordingly. By contrast, we assume it is too costly for the consumer to react in real time to new information about prices. This creates 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 publicly observed, demand $D^s(\overline{p})$ obtains from the maximization of the net surplus

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

This demand function is decreasing with the retail price $\overline{p}$, with a derivative $D_p^s < 0$. We also define the elasticity of demand

$$\varepsilon^s(p) = -\frac{pD_p^s(p)}{D^s(p)} > 0. \tag{2}$$

Disconnection, and the Value Of Lost Load. For technological reasons, it is impossible to partially disconnect a consumer: either a consumer sees his demand satisfied, or he is disconnected and gets nothing. In this case, the consumer obtains a gross surplus $v^s(0)$, and pays only the fixed fee A.

Now, suppose that global production 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 $v^s(e) - v^s(0)$. The associated aggregate social loss is called the Value of Lost Load (VOLL), and equals

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

Note that ℓ^s is decreasing with e, by concavity of the surplus function v^s. It measures the social loss associated to a reduction by one unit of electricity supply, when all consumers are connected and face the same retail price, and when our assumption of rigid demand management holds.[16] We shall assume that in any case ℓ is high enough to lie above any conceivable marginal cost of production.

Rolling and systemic blackouts. The wholesale electricity market must be balanced at any time, and this is an impossible task when demand exceeds the available capacity, or when the grid is impaired, say by a storm or a fire. In such cases, the rigidity of demand imposes that some consumers be cut off from the grid. But disconnecting some consumers sometimes implies disconnecting simultaneously some of the production units. This also re-routs electricity towards transmission lines that may already be congested, possibly triggering additional disconnections. The outcome may then be either a rolling blackout (where some consumers are disconnected but all available capacity remains connected), or a systemic blackout (during which entire parts of the grid are out of function, so that some available capacity is off-line).[17]

To model blackouts in a simple but general manner, we assume that the type and the size of a blackout are a function of both the state of the world s, and of the following

[16]Note that, if demand could be flexibly managed, then the VOLL would be $v'(e)$, which is lower than $\ell(e)$ by concavity of v. 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.

[17]For instance, in February 2021, many of the natural gas processing facilities in Texas were curtailed, resulting in an almost 50 percent decline in natural gas production, and in a significant amount of natural-gas fired generating capacity being made unavailable. See Wolak (2022) for a discussion.

capacity index:

$$\kappa = \frac{K}{D}. \tag{3}$$

The capacity index κ is thus the ratio of available capacity to realized demand. As such, it measures the robustness of the grid to random shocks on supply and demand. It presents the advantage of neutralizing scale effects. We moreover assume that disconnection occurs uniformly, both for production units and for consumers, so that disconnection affects only the size of each side of the market, and not its composition. As a consequence, the proportion of connected production units is simply a function

$$n^s(\kappa) \in [0, 1].$$

This function could be estimated from aggregate data on demand, capacity, and connections. For the moment, we only assume it is weakly increasing in κ, and we define the elasticity of the proportion of connected producers to the capacity index as

$$\nu^s(\kappa) = \frac{\kappa \frac{\partial n^s}{\partial \kappa}(\kappa)}{n^s(\kappa)}. \tag{4}$$

Note that this elasticity is zero when all producers are connected ($n = 1$), or when producers' disconnections are purely exogenous. ν thus measures the endogeneity of blackouts.

Concerning the demand side, we assume that the grid is managed so as to serve as many consumers as possible. As discussed in the Introduction, this assumption seems reasonable, and in line with actual practice.[18] Since the connected capacity is nK and each consumer demands D, we obtain that the proportion m of connected consumers is exactly

$$m^s(\kappa) \equiv \min(\frac{nK}{D}, 1) = \min(n^s(\kappa)\kappa, 1). \tag{5}$$

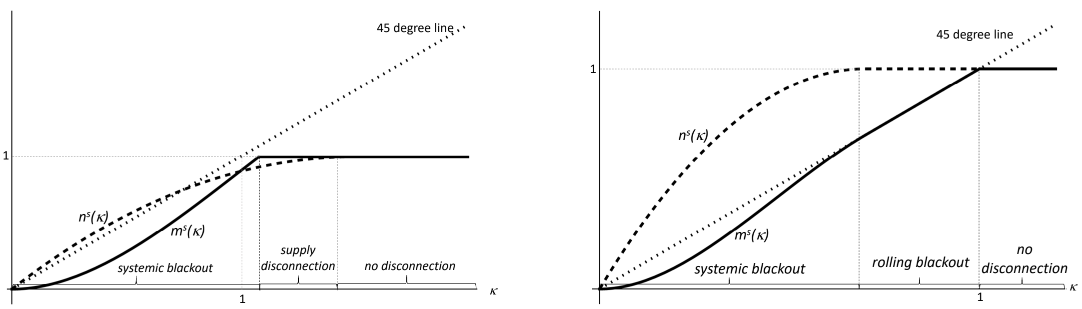
This framework encompasses a variety of possible regimes, which we describe in Table 1. When $m = 1$ (left column), supply is sufficient to satisfy demand, whereas in the right column ($m < 1$), the lack of capacity leads to blackouts. In the top-left box ("no disconnection"), all producers and consumers are connected to the grid, and connected capacity is sufficient to ensure that all demands are satisfied. From this regime, an

[18]In fact, in some cases it could be welfare-improving to disconnect more consumers than is needed, so as to make producers compete for the remaining demand. We leave this possibility for future research.

	$m^s(\kappa) = 1 \leq n^s(\kappa)\kappa$	$m^s(\kappa) = n^s(\kappa)\kappa < 1$
$n^s(\kappa) = 1$	**no disconnection** $K > D$	**rolling blackout** $K = mD$
$n^s(\kappa) < 1$	**supply disconnection** $nK > D$	**systemic blackout** $nK = mD$

Table 1: m is the proportion of fully served consumers, n is the proportion of connected producers.

incident on the grid (captured by the state of nature s), or a surge in demand, or a reduction in capacity (both impacting the capacity index κ), may impact the proportion n of connected producers and the proportion m of served consumers. If the capacity index remains high enough, some producers may be disconnected while demand is still fully served (bottom left, "supply disconnection"). Alternatively, if n remains high while κ drops, some consumers have to be disconnected even though producers remain connected ("rolling blackout"). Which path realizes thus depends on whether n or κ bears the brunt of the shock, as illustrated in Figures 1(a) and 1(b). Finally, more major shocks may impair the functioning of entire parts of the grid, and lead to a "systemic blackout" in which both supply and demand are disconnected.



(a) Producers disconnected first.

(b) Consumers disconnected first.

Figure 1: Two possible paths as κ is reduced, for a given state s.

Note finally that thanks to the uniform disconnection assumption, each producer can be seen as endowed with the aggregate cost function C, and each of the n connected producers has to satisfy the demand of m/n consumers. Therefore, the realized total cost equals

$$nC^s\left(\frac{m}{n}D^s(\overline{p}), X\right).$$

3 Optima

In this Section, we characterize the surplus-maximizing allocations, specifying the ex-ante investments, and the productions and consumptions in each state. Consumptions are constrained by the rigidity of demand management: there must exist a retail price $\overline{p}$ such that in each state s consumption is $D^s(\overline{p})$, or zero if the consumer is disconnected. From Definition (3), the value $\kappa(s)$ of κ in state s is such that

$$\kappa(s)D^s(\overline{p}) = \sum_k x_k K_k^s \equiv K^s(X). \tag{6}$$

The consumers' gross surplus in state s is

$$m^s(\kappa(s))v^s(D^s(\overline{p})) + (1 - m^s(\kappa(s)))v^s(0),$$

from which we have to subtract the production cost

$$n^s(\kappa(s))C^s\left(\frac{m^s(\kappa(s))}{n^s(\kappa(s))}D^s(\overline{p}), X\right)$$

and the investment cost

$$\sum_k I_k(x_k).$$

Optimal allocations maximize the expectation over s of this aggregate surplus with respect to the capacity indexes $\kappa(s)$ in each state s, the retail price $\overline{p}$, and the investments X, under Constraint (6). Since the proportion of connected producers is measured by n, we associate to the constraint in state s a multiplier $n^s(\kappa(s))\beta(s)$. Therefore, $\beta(s)$ measures the social value of one additional *connected* unit of capacity in state s, and we expect it to be nonnegative.

Let us first provide necessary conditions associated to the choice of $\kappa(s)$ (some details of the derivations are gathered in the Appendix). Note that for the sake of clarity, we often omit arguments in the formulas below. In states where all consumers are served, i.e. when $m = 1 \leq n\kappa$ as stated in Definition (5), the objective reduces to

$$v(D) - nC\left(\frac{D}{n}, X\right),$$

to be maximized under Constraint (6). From the Definition (4) of the elasticity of n , we obtain

$$\text{If } m = 1, \qquad \beta(s) = \frac{\nu}{\kappa n}\Big[C_Q(\frac{D}{n}) - AC(\frac{D}{n})\Big]. \tag{7}$$

Here, the only value of an additional unit of capacity is to increase the capacity index and finally the number of connected producers, with the elasticity ν. This in turn reduces the production cost $nC(\frac{D}{n})$, to the extent that marginal cost is above the average cost. This effect is zero if all producers are connected (the no-disconnection regime in Table 1), or if disconnection of producers is purely exogenous ($\nu = 0$), and it is strictly positive otherwise (supply disconnection in Table 1).

The picture is more complex for states in which a blackout occurs , i.e. $m = n\kappa < 1$, from Definition (5). Then the objective reduces to

$$n\kappa(v(D) - v(0)) - nC(\kappa D, X),$$

to be maximized under Constraint (6). After some simplifications, we obtain

$$\text{If } m < 1, \qquad \beta(s) = \ell(D) - C_Q(K) + \nu(\ell(D) - AC(K)). \tag{8}$$

Now, the additional unit of capacity directly increases the proportion m of connected consumers, with a gain equal to the VOLL, at the price of a marginal increase in the cost (regime rolling blackout in Table 1). It also has an indirect effect, when some producers n are disconnected: then a higher capacity index increases the number of connected producers n, and this allows to produce more and to serve more consumers, now at the average cost since the additional connected producers are drawn at random in the population of producers. This second effect corresponds to a contribution to a public good, i.e. the stability of the grid. It is scaled by the value of the elasticity ν, which measures the endogeneity of blackouts. Finally, the social value β of additional capacity is positive since the VOLL ℓ is assumed to be high enough compared to costs.

Regarding investments, the first-order condition with respect to the number x_k of units of type k can be simplified using Property (1) to become (assuming it is optimal to use at least one such unit):

$$I'_k(x_k) = E\Big[n\pi_k(C_Q)\Big] + E[\beta n K_k]. \tag{9}$$

The two first terms measure the private incentives to invest, assuming electricity is priced at the global marginal cost (though we shall soon explain why this is unlikely to be the case in equilibrium). The last term is the expected social value of this additional unit.

We finally turn to the first-order condition with respect to the retail price $\overline{p}$, in which we use Definition (2) of the elasticity of demand to get:

$$E\Big[\varepsilon m D(\overline{p} - C_Q)\Big] = E[\varepsilon \beta n K]. \tag{10}$$

This means that the optimal retail price should equal an average of marginal costs, computed with a complex distribution that takes into account both total consumption mD and the elasticity of demand ε, to which must be added a term which measures the beneficial impact of a reduction in demand on the value of the capacity index.

4 Competitive Equilibria (with price cap and rationing)

When applied to wholesale electricity markets, the standard notion of competitive equilibria faces a difficulty: because retail pricing is rigid, no price can balance demand and supply when available capacity is less than realized demand, so that the idea of price-taking agents becomes awkward. The notion of competitive equilibria we shall use thus has to be augmented with regulatory interventions.

The first regulatory reaction to an imbalance on the market is to disconnect some consumers. But when the minimum number of consumers is disconnected, as we have assumed, each active producer remains indispensable to the satisfaction of the remaining consumers. With non-infinitesimal producers, being indispensable means being able to increase prices without bounds. When observing such a surge in prices, the second regulatory reaction consists in imposing a price cap P, set ex-ante at a reasonable level to settle transactions between producers and retailers.

In our model, we introduce these two regulatory interventions as follows. In states where there is excess capacity, producers act as price-takers, and electricity is priced at marginal cost, so that the wholesale price of electricity is:

$$\text{If } m = 1, \text{ then } p(s) = C_Q^s\left(\frac{D}{n}, X\right). \tag{11}$$

Otherwise, when demand is rationed active producers become indispensable and can extract an arbitrarily high price for their production. In such a case, we require that the electricity price be capped at P, assumed high enough compared to costs so as to ensure that all available capacity is put to use:

$$\text{If } m < 1, \text{ then } p(s) = P > C_Q^s(K^s(X), X). \tag{12}$$

This modeling choice can be seen as a simple manner to model imperfect competition. [19] As discussed in the Introduction, a natural candidate for the cap is the VOLL, but various difficulties typically lead regulators to set the price cap below the VOLL. This is first to reduce the risk that retailers bear by buying at the risky wholesale price to sell at the fixed retail price. This is also to reduce the market power enjoyed by dominant producers (see Wolak (2013), and Fabra (2018), among many others). We shall thus take P as exogenously given, plausibly below the VOLL, but definitely above production costs so as to ensure that all available capacity is indeed used in case of blackouts, as stated in (12).

Still, setting the price cap below the VOLL creates two additional distortions, requiring two additional instruments. The well-known "missing money" problem underlines that with a low cap, producers are not rewarded sufficiently for their investments. One should thus introduce additional incentives for capacity creation. We therefore introduce a subsidy σ_k for each unit of type k. In addition, a low cap also reduces the price retailers

[19] Interestingly, it introduces a discontinuity in the wholesale price between states of the world in which $m < 1$ (blackouts) and states in which $m = 1$ (no blackouts). This discontinuity would still appear if one were to model more precisely imperfect competition between producers. In this alternative model, there would appear a threshold for m such that, when m is above this threshold, then pricing would obey the standard rules of, say, Cournot pricing, as for example in Elliott (2024); while if m is below the threshold, strategic producers would withdraw capacities so as to trigger the imposition of the price cap P. Whether such opportunistic behaviors may be discouraged by the fear of public reactions is difficult to assess; McRae and Wolak (2019) claim to have identified such behaviors for the Colombian electricity market. The point here is that this discontinuity is not pathological: it is a natural consequence of the possibility of blackouts, and of strategic behavior on the supply side.

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.

This being clarified, let us now turn to the derivation of competitive equilibria, for given values of the price cap P, the retail tax τ, and the various subsidies to capacity σ. As before, one has to find an allocation in each state of nature, together with endogenous equilibrium prices which are the rigid retail price $\overline{p}$ and the flexible wholesale prices $p(s)$. Constraint (6) defining $\kappa(s)$ still holds, by definition of the variables involved. Wholesale prices $p(s)$ are determined in each state, as stated in (11) and (12).

Recall that competitive retailers buy electricity at the wholesale price $p(s)$, to resell it to consumers at the fixed retail price $\overline{p}$, including the retail tax. The retail contract also specifies a fixed fee A. Competition should then select the contract $(A, \overline{p})$ that maximizes the consumer's surplus

$$E\Big[m\Big(v^s(D^s(\overline{p})) - \overline{p}D^s(\overline{p})\Big) + (1-m)v^s(0)\Big] - A,$$

under the constraint that profits be nonnegative:

$$A + (\overline{p} - \tau)E[mD^s(\overline{p})] \geq E[p(s)mD^s(\overline{p})]. \tag{13}$$

This constraint clearly binds, so that the equilibrium contract gives zero-profit to retailers and maximizes over $\overline{p}$ the joint surplus

$$E\Big[m\Big(v^s(D^s(\overline{p})) - (p(s)+\tau)D^s(\overline{p})\Big) + (1-m)v(0,s)\Big].$$

This result appears also in Joskow and Tirole (2006, 2007). These works also underline that retail contracts raise difficulties of their own when consumers are heterogenous, since different consumers with different load profiles entail different expected costs for the retailer.[20] For the sake of simplicity, we abstract from these difficulties, and the following first-order condition characterizes the equilibrium retail price:

$$(\overline{p} - \tau)E[\varepsilon mD] = E[\varepsilon mDp(s)]. \tag{14}$$

[20]This implies in particular that the use of such contracts creates an adverse selection problem for the retailers, as the profits made on each consumer depend on its individual load profile–*i.e.*, on whether they consume electricity at peak hours or not. We ignore this problem in this article, and we refer to Joskow and Tirole (2006) for a related study, and to Leslie, Pourkhanali and Roger (2021) for an estimate of cross-subsidies between consumers who sign the same contract.

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: for all k, one has

$$I'_k(x_k) = E[n\pi_k^s(p(s))] + \sigma_k. \tag{15}$$

Overall, a competitive equilibrium thus has to verify the first-order conditions in (11)-(12)-(14)-(15).

5 Decentralization

The next step is to determine the values of regulatory instruments that decentralize a given optimal allocation, for a given price cap P. To state the result more clearly, we define in each state the following measure of missing incentives:

$$\Delta(s) = \begin{cases} \ell(D) - P + \nu\left(\ell(D) - AC(K)\right) & \text{if } m < 1 \quad \text{(16a)} \\ \dfrac{\nu D}{nK}\left(C_Q(\dfrac{D}{n}) - AC(\dfrac{D}{n})\right) & \text{if } m = 1. \quad \text{(16b)} \end{cases}$$

Without further ado, let us state our main result:

Proposition 1 *Let the price cap P be given. To decentralize a given optimal allocation, one needs to set the retail tax τ and the capacity subsidies σ_k as follows:*

$$\tau E[\varepsilon m D] = E[\varepsilon \Delta n K] \tag{17}$$

$$\textit{For all } k, \qquad \sigma_k = E[\Delta n K_k]. \tag{18}$$

The proof of this result relies on comparing the first-order conditions (7)-(8)-(9)-(10) characterizing an optimum, to the first-order conditions (11)-(12)-(14)-(15) characterizing an equilibrium. The resulting values for the tax and the subsidies thus constitute only necessary conditions. Nevertheless, it is remarkable that this simple set of regulatory instruments is able to fully reconcile these first-order conditions.

The general structure of the result is also interesting. Our measure of missing incentives is the sum of two terms, one for the missing money problem, which is present in Equation (16a) when the price cap is active, and a second one for the public good problem, weighted by ν. Since both problems require to reduce demand and increase investment, the measure of missing incentives appears in the formulas for both the retail tax and the capacity subsidies. Observe also that Δ relates directly to the social value of connected capacity β defined in Equations (7) and (8) in our study of optima: indeed, we have

$$\Delta = \beta - 1_{m<1}(P - C_Q(K)), \tag{19}$$

where the second term appears because in the competitive equilibrium the price cap provides part of the missing incentives. These are nice normative results, which complement the positive analysis in Joskow and Tirole (2007). We now list a few qualitative properties of the optimal regulation.

A retail tax is needed. Indeed, the optimal retail tax is zero only when the right-hand side of (17) vanishes. It is easily shown that this requires that the price cap be set at a value which is above an expectation of the VOLL, conditional on demand rationing.[21] These are states when the VOLL is likely to be high. Recall also that existing regulations typically set a cap below the VOLL. The model thus supports the creation of a positive retail tax, simply because by reducing demand it alleviates both the missing money problem and the public good problem.

Subsidies are for capacities. For each production unit, it is important to distinguish the nominal capacity, which is a characteristic independent from the state of nature, from available capacities in each state (K_k^s in the model), and finally from connected capacities in each state (nK_k^s in the model). In our model, what matters for the stability of the grid are the available capacities K_k^s, and consequently they should be subsidized. Notice also that the value of Δ in (16b) depends on the connected capacities nK_k^s.[22]

[21] A proof is as follows. The optimal retail tax decreases with the price cap, with a slope between -1 and 0, because from (17) we have $\frac{\partial \tau}{\partial P} E[\varepsilon m D] = -E[1_{m<1}\varepsilon n K] = -E[1_{m<1}\varepsilon m D] \geq -E[\varepsilon m D]$. The retail tax is thus zero exactly when P satisfies the following equation, in which we have used the fact that $mD = nK$ when $m < 1$:

$$E[1_{m<1}\varepsilon n K(P - \ell(D))] = E\Big[\nu \varepsilon m D\Big(1_{m<1}\ell(D) + 1_{m=1}C_Q(\frac{D}{n}) - AC\Big)\Big].$$

Since the right-hand side is nonnegative, this shows that the cap needs to be at least an expectation of the VOLL, conditional to rationing.

[22] In actual regulatory systems, information about these values is typically requested in advance from producers, who are penalized if their reports turn out to be inflated in case these capacities are called

By contrast, some regulatory systems operate directly on wholesale prices. The Operating Reserve Demand Curve system used in the Texas ERCOT adds a premium to the wholesale price when reserves fall below a given threshold. This increases the financial revenues from production, and thus the incentives for investing ex-ante; but one may as well argue that in the short-run producers are rewarded for the scarcity of reserves, and that market power becomes even more valuable. Such distortions to market pricing are not justified in our model.

Alongside these capacity mechanisms, actual regulatory systems also include the management of reserves, i.e. some idle capacity that can be made active very rapidly. Siuch reserves are costly, and they are sometimes imposed, sometimes allocated through auctions. Once more, the key issue is whether such systems distort market pricing, by subsidizing producers for reducing production.[23]

Optimal subsidies to capacity depend on intermittency. Let's begin with the simplest case: when there is no intermittency, then all units of capacities play the same role for the stability of the grid, and therefore should receive the same subsidy. One can extend this result to stochastic and independent disconnections, using a law of large numbers; or to the case when all capacities are subject to the same random multiplicative shock.[24] But these are very restrictive cases. In general, and especially because intermittent sources rely on weather variables that are quite volatile, optimal subsidies to capacity cannot be uniform. The computation of the expectation in Equation (18) seems quite complex, as it takes into account the correlation between the missing incentives and the connected capacity. As explained above, one can still compute Δ in each state, and use it to allocate subsidies. All production plants are treated symmetrically, with the same value of Δ; but this value applies to connected capacities that are much more volatile for intermittent sources. Intermittent sources that produce only when Δ is low should thus

upon.

[23]See Borenstein, Bushnell and Mansur (2023) for a discussion of regulatory systems.

[24]For this latter case, suppose available capacities equal nominal capacities, multiplied by a common shock h: $K_k^s = h(s)\overline{K}_k$. Then Equation (18) characterizes an optimal subsidy per unit of nominal capacity which is indeed independent from the unit type:

$$\frac{\sigma_k}{\overline{K}_k} = E[\Delta n h(s)].$$

receive much lower subsidies. In particular, when demand is relatively small and the weather is favorable, then production can be fully ensured by renewables, whose marginal cost is a constant. Then marginal cost equal average cost, and in this case Δ is zero, as seen in Formula (16b). In other words, these states should not contribute at all to increase the subsidies to renewable capacities.

Exogenous blackouts, and the energy-only paradigm. This paradigm proposes to rely only on scarcity pricing to regulate investments. To discuss it, assume firstthat producers may be disconnected only for exogenous reasons, unrelated to possible grid imbalances. Then the elasticity ν of the proportion of connected producers to the capacity index is nil, and the definition of Δ reduces to

$$\Delta(s) = \ell^s(D^s) - P \text{ if } m < 1, \text{ and zero otherwise.}$$

Assume moreover that the VOLL $\ell^s(D^s(\overline{p}))$ is independent from s. Then missing incentives are identically zero if the price cap is set equal to the VOLL, so that both the retail tax and the capacity subsidies can be set to zero, according to Proposition 1. This vindicates the energy-only paradigm: setting a price cap equal to the VOLL is sufficient. But this result relies on a number of assumptions that we list below.

Firstly, as already noticed in this article there are good reasons for setting the price cap below the VOLL: one has to protect retailers from facing high wholesale prices, and a lower price cap also reduces the exercise of market power by producers.

Secondly, the VOLL might well depend on the state of nature. Then Δ becomes a non-zero random variable, and Proposition 1 indicates that setting a price cap equal to some expected value of the VOLL is not enough: the other regulatory instruments are still needed, depending on the correlation between Δ and the various capacities.

Finally, the assumption that blackouts are exogenous does not seem reasonable. What we observe in the case of the Texas blackout, or in the more recent case of Spain, is that blackouts are more likely in periods of high demand and low availability of capacity, so that ν is not zero. The energy-only paradigm seems to be based on demanding assumptions.

The case of a multiplicative shock on demand. In addition, we are able to obtain striking results in a particular case often used in the theoretical literature, as well as in empirical studies of the electricity sector.[25] Suppose that demand is impacted by a multiplicative shock b, so that:

$$D(p, s) = b(s)d(p), \tag{20}$$

where $b(s)$ is positive and the function d is decreasing. Under this assumption, a number of straightforward results follow (see Section C in the Appendix for details of the derivations):

Proposition 2 *In the case of a multiplicative shock on demand,*

i) The competitive access fee A is zero;

ii) The elasticity of demand $\varepsilon^s(\overline{p})$ and the Value of Lost Load $\ell^s(D^s(\overline{p}))$ do not depend on s;

iii) The optimal regulation is such that the revenues from the retail tax exactly balance the cost of subsidies:

$$\tau E[mD] = \sum_k x_k \sigma_k.$$

The assumption of a multiplicative shock does not seem very restrictive in practice, and it allows to give a clear meaning to the VOLL. Remarkably, it also yields the result that the regulatory budget is exactly balanced. This remarkable identity is new in the literature, and may be useful as a guide for regulatory policies.

6 Taxing CO_2: effects on supply and demand

Our model can also be used to give insights on how a tax on greenhouse gases emissions impacts all components of the electricity market. In most studies, the impact of CO_2 taxation is studied only through the lens of its impact on supply: production from fossil fuels is reduced, with a substitution towards clean substitutes such as nuclear or renewables. As discussed in the Introduction, an exception is Holland, Mansur and Yates (2022), which offers a model of the electricity sector calibrated on US data, and which explicitly considers the electrification of uses. But their model does not allow for blackouts, and carbon

[25]Even recent papers such as Elliott (2024) typically assume a multiplicative shock and an iso-elastic demand function. This last assumption is in fact not needed for Proposition 2 to hold.

taxation is assumed to bear only on the supply side; in that paper, the electrification of uses is thus exogenous. By contrast, we think it important to also include the impact of carbon taxation on demand. Hence, in our model a carbon tax modifies simultaneously supply and demand, with consequences on the reliability of the grid, and finally on the optimal regulation.

For simplicity, fossil fuels are introduced under the general label of gas. Theoretical results are complemented by simulations; let us emphasize from the start that these simulation exercises have no claim toward realism.

6.1 Extending the model

We introduce a Pigovian carbon tax t in our model, equal to the constant marginal damage of greenhouse gases emissions. This is clearly for the sake of simplicity: with an optimal tax, fighting climate change does not require to distort regulatory instruments or to create new ones, such as subsidies to renewables. The only consequence in our model is that the price of gas is linearly increased, from a fixed world price c_g to $p_g = c_g + t$.

The effects on the supply side are easily handled. Assume that gas is the only input used in a technology k to produce electricity at constant marginal cost: this constant marginal cost then equals the after-tax gas price $p_g = c_g + t$. To take into account changes in the gas price, we therefore re-define the cost function into $C(Q, X; p_g)$.

On the demand side, one expects consumers to react to an increase in the price of gas by substituting electricity to gas or fuel, for heating or transportation. Suppose that in each state s the representative consumer can now consume electricity e and gas g to satisfy their needs for energy, yielding a surplus

$$\mu^s S\left(\frac{e}{\mu^s}, \frac{g}{\mu^s}\right)$$

where μ^s is the importance of energy in state s. A higher value for μ thus increases the need for both electricity and gas. The surplus function S is assumed strictly concave with a negative cross-derivative ($S_{12} < 0$), so that gas and electricity are substitutes. One can then re-define the indirect utility function

$$v^s(e; p_g) \equiv \max_g \left[\mu^s S\left(\frac{e}{\mu^s}, \frac{g}{\mu^s}\right) - p_g g\right]$$

and the VOLL function

$$\ell^s(e;p_g) \equiv \frac{v^s(e;p_g) - v^s(0;p_g)}{e}.$$

One advantage of our particular surplus function is that it preserves the property of multiplicative shocks on demand: indeed, by maximizing

$$\mu^s S\left(\frac{e}{\mu^s}, \frac{g}{\mu^s}\right) - \overline{p}e - p_g g$$

we get

$$D^s(\overline{p}, p_g) = \mu^s d(\overline{p}, p_g), \tag{21}$$

where d is decreasing in $\overline{p}$ and increasing in p_g. The results in Proposition 2 thus still hold in this extended model. In particular, the value of the VOLL at the optimum

$$l^s(D^s(\overline{p}, p_g); p_g)$$

does not depend on the state of nature.

6.2 The scale experiment

A simple way to analyze the effects of carbon taxation is as follows. Thanks to the separability property in Equation (21), one can see that an increase in the price of gas from p_g to p'_g causes demand to be multiplied by a constant factor that does not depend on the state of nature:

$$\theta \equiv \frac{d(\overline{p}, p'_g)}{d(\overline{p}, p_g)} > 1.$$

This motivates the following scale experiment: faced with this uniform increase in electricity demand, one could simply multiply all investments $X = (x_k)$ by the same factor θ, keeping the same production plan (q_k^s) for each production unit of each type, and leaving the retail price $\overline{p}$ unchanged. Then the ratio κ of capacity to demand would be unaffected, leaving the proportions m and n unchanged. In the absence of other effects, one would not need to modify the regulation, provided it was initially optimal.

This simple reasoning neglects however three specific effects of carbon taxation, which we now discuss. Recall that total surplus equals

$$E\Big[mv^s(D;p_g) + (1-m)v^s(0;p_g) - nC(\frac{m}{n}D, X; p_g)\Big] - \sum_k I_k(x_k).$$

1. A direct effect is that producing electricity from gas becomes more costly. This should lead producers to substitute nuclear or renewables (or both) to production from fossil fuels.

2. A less-studied effect is that total investment costs $\sum_k I_k(x_k)$ may increase non-linearly if some of the functions I_k are strictly convex. This is likely to be the case for renewables, because productive spots for these technologies are scarce, or located too far away from consumers. Therefore, we expect the quantities x_k for renewables to increase by less than θ, while nuclear capacities could increase by more.

3. Finally, an increase in p_g impacts demand of both gas and electricity. To study this effect, note that the consumer surplus

$$m^s(\kappa(s))v^s(D^s(\overline{p}, p_g), p_g) + (1 - m^s(\kappa(s)))v^s(0, p_g)$$

equals

$$v^s(0; p_g) + m^s(\kappa(s))D^s(\overline{p}, p_g)\ell^s(D^s(\overline{p}, p_g); p_g).$$

The first term is a direct effect: the availability of gas allows consumers to hedge against the risk of a blackout. Therefore, the movement towards the electrification of uses makes consumers more vulnerable to blackouts, since their outside options are reduced. This represents a net loss to consumers, but this loss is fixed, and *per se* it does not support any change in regulatory decisions.

The second term is the product of three terms: the probability of being connected to the grid (which in this scale experiment has not changed), demand (which is multiplied by θ), and the VOLL. To study the effect of carbon taxation on this last term, let us rewrite the VOLL as

$$\ell^s(D^s(\overline{p}, p_g); p_g) = \frac{v^s(D^s; p_g) - v^s(0; p_g)}{D^s} = \frac{\int_{\overline{p}}^{+\infty} \rho D_p^s(\rho, p_g) d\rho}{\int_{\overline{p}}^{+\infty} D_p^s(\rho, p_g) d\rho}.$$

Since $\overline{p}$ has not changed in this scale experiment, we see that the VOLL may increase or decrease with respect to the price of gas. In fact, with a log-linear demand

$$D^s(\overline{p}, p_g) = \mu^s d_e(\overline{p}) d_g(p_g),$$

we obtain that the VOLL does not depend on the gas price, so that we may as well consider that this term is simply rescaled by the factor θ.

To summarize, an increase in the price of gas due to carbon taxation leads to an increase in the demand for electricity. The scale experiment attempts to handle this electrification of uses through simple changes in scale. Using this lens, we distinguish two categories of effects. Firstly, there are effects that directly impact total surplus. Blackouts have more detrimental effects, because consumers cannot rely on cheap gas anymore. Producers using fossil fuels bear higher costs. Renewables face decreasing returns to scale, due to the scarcity of favorable locations. Secondly, these effects in turn lead to adjustments in decisions, compared to a simple change in scale. These distortions are difficult to sign on the demand side; in fact, they vanish under the assumption of a log-linear demand. By contrast, they are clear on the supply side: expected gas consumption must decrease, and one expects a more-than-proportional development of nuclear, and a less-than-proportional development in renewables.

In addition, higher production costs should lead to an increase in the retail price, which in turn reduces demand. As a result, the net effect on the capacity index is ambiguous—it may increase or decrease—and the implications for blackout risk are similarly uncertain. To address these analytical ambiguities, in the following section we turn to numerical simulations. These simulations enable us to determine the sign of the impact of a carbon tax on all relevant variables, even when analytical methods are inconclusive, and also to assess the quantitative magnitude of these effects.

6.3 Simulations

We simulate a simple economy in which electricity can be produced using two technologies: renewables ($k = 1$, with zero marginal cost) and gas ($k = 2$). As before, the gas price p_g includes a carbon tax t, and we examine the effects of varying p_g. We distinguish between two scenarios. In Scenario S (Supply only), we assume that producers bear the full cost of p_g, while consumers face an unchanged retail gas price. In Scenario DS (Demand and Supply), the gas price p_g applies to all agents, both producers and consumers. This framework allows us to disentangle the separate effects of carbon taxation on the supply and demand sides of the market.

Demand is affected by a multiplicative shock s^D, while the production capacity of renewables is affected by a multiplicative shock s^S. Hence, the state of nature s is a pair (s^D, s^S), and for simplicity we assume that these shocks are uniform and independent. The consumers' surplus displays a constant elasticity of substitution:

$$s^D S\left(\frac{e}{s^D}, \frac{g}{s^D}\right) = a s^D \left(b\left(\frac{e}{s^D}\right)^\rho + c\left(\frac{g}{s^D}\right)^\rho\right)^{\frac{\gamma}{\rho}}$$

with a, b and c positive and $0 < \gamma < \rho < 1$. These inequalities ensure that S is concave, and that the two goods are substitute in each state of nature.

Once more for simplicity, the proportion $n^s(\kappa)$ connected production units is independent from s, and it is given an increasing and concave shape:

$$n(\kappa) = \begin{cases} \frac{\kappa}{\overline{\kappa}^2}(2\overline{\kappa} - \kappa) & \text{for } \kappa \leq \overline{\kappa} \\ 1 & \text{for } \kappa > \overline{\kappa} \end{cases}$$

Appendix D details the various parameters used in the simulations, including the parameter $\overline{\kappa}$. We maximize total surplus numerically, while varying the gas price p_g. Figure 2 shows how this increase in price reduces total surplus, and especially so when the effect on demand enters the picture. As we shall see, this difference between scenario S and scenario DS vanishes when considering the agents' decisions, in line with a remark made in the previous subsection.

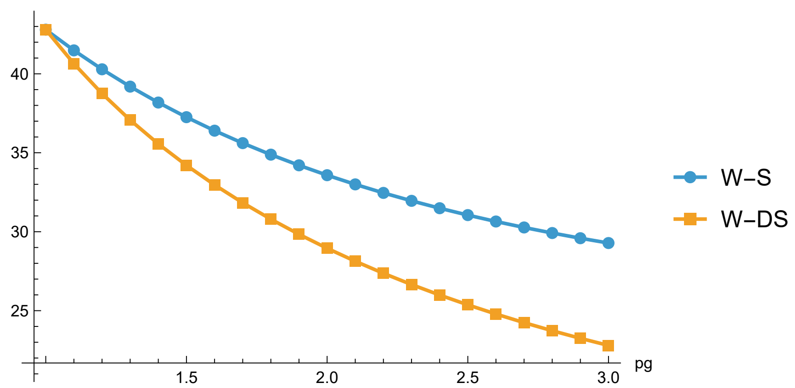
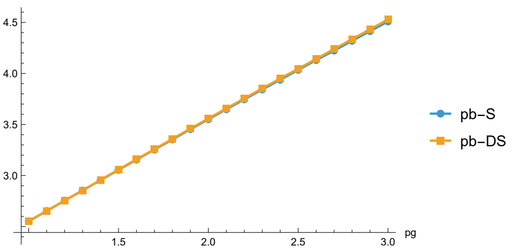
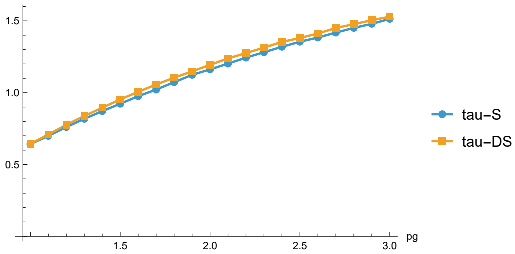


Figure 2: Total surplus as a function of the carbon tax.

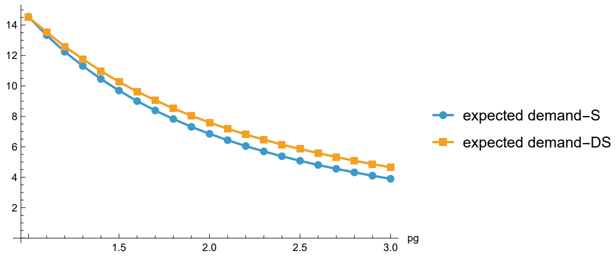
Figure 3 focuses on the demand side. The retail price goes up with the price of gas, because production costs are higher, but in fact mainly because the optimal retail tax increases significantly. These effects are strong enough to reduce total demand, overcoming the substitution effect in demand, which should support a higher electricity demand.



(a) Retail price.



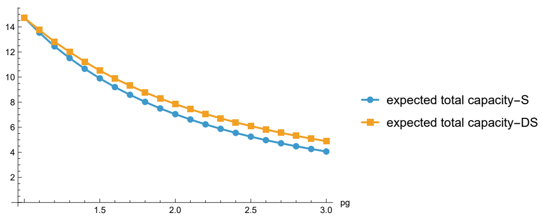
(b) Retail Tax.



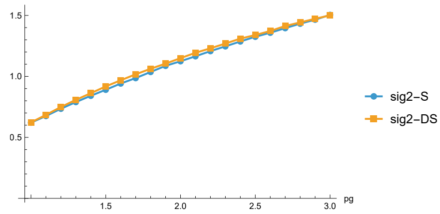
(c) Expected demand.

Figure 3: Retail price, expected demand, and retail tax, as a function of the carbon tax.

We then need to understand why the retail tax increases so much. In Figure 4, we see that the expected total capacity goes down with the gas price, roughly in the same proportion as demand. But the capacity index is on average much higher: at the highest value for the gas price, on average available capacity exceeds demand by half. Simultaneously, we see in Figure 5 that the variance of this index also increases; and in Figure 6, we observe that the probability of blackouts goes down.



(a) Expected total capacity.



(b) Expected Capacity index.

Figure 4: Total capacity and capacity index as a function of the carbon tax.

The interpretation we propose is the following. A higher gas price makes producers switch to renewables, as observed in Figure 7. But because of intermittency, the variance of the capacity index increases: in favorable states, it is windy or sunny, and total capacity is high compared to demand; in unfavorable states, only gas plants are active, and it

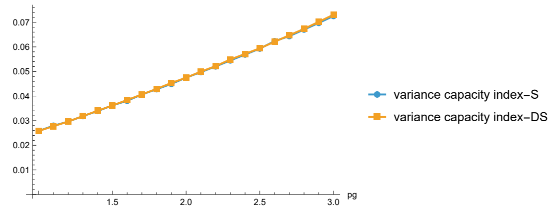
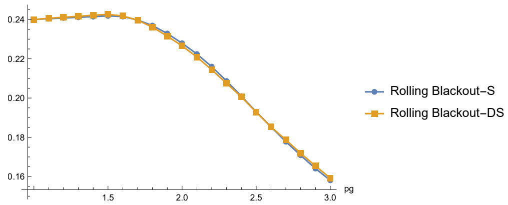
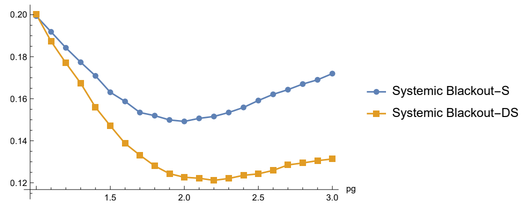


Figure 5: Variance of the capacity index, as a function of the carbon tax.



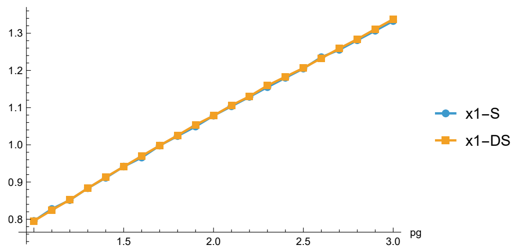
(a) Probability of a rolling blackout.



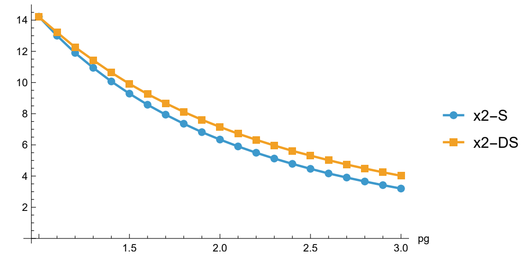
(b) Probability of a systemic blackout.

Figure 6: Probability of a blackout as a function of the carbon tax.

becomes more difficult to stabilize the grid. Because consumers have no outside option, blackouts are more detrimental to consumers. It is thus important to increase the capacity index in unfavorable states. This can be done by increasing the retail tax, so as to reduce demand. Quite paradoxically, this also motivates an increase in the subsidies to producers who use gas, since they are badly needed in states with unfavorable weather, as can be seen in Figure 8. An optimal policy may thus simultaneously tax greenhouse gases emissions, and subsidize electricity production from fossil fuels.

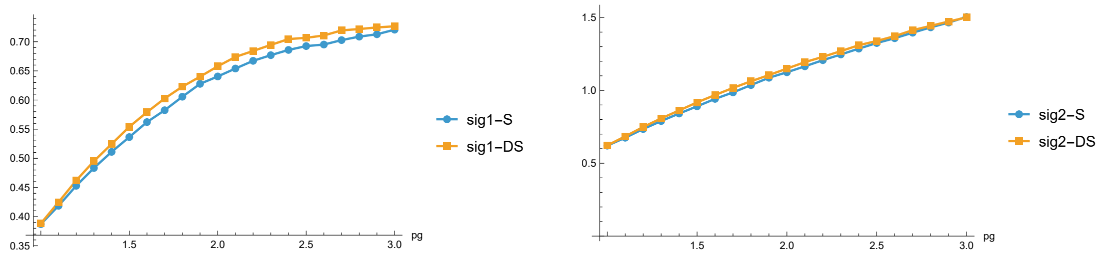


(a) Investment in renewables.



(b) Investment in gas.

Figure 7: Investments as a function of the carbon tax.



(a) Subsidies to renewables.

(b) Subsidies to gas.

Figure 8: Capacity subsidies as a function of the carbon tax.

7 Conclusion

We have set up a stylized but general model of the electricity sector. We have focused on two issues: the inability to balance the market creates the need for rationing and for introducing a price cap, which in turn creates a missing money problem; and the public good problem, associated to the reliability of the grid, which depends on the excess capacity relative to demand. We have shown that these two issues can be tackled efficiently thanks to a retail tax, and to differentiated subsidies to investment. Under the assumption that demand is only affected by a multiplicative shock, we have established that it is optimal to balance the budget of the regulatory agency.

We also have provided precise formulas in order to compute optimal subsidies to capacities. These formulas essentially measure the social values of each capacity, both when providing electricity and when contributing to the stability of the grid. Without surprise, they depend a lot on when electricity is produced: subsidies should be higher for capacities that produce during peaks, and that are flexible enough to stabilize the grid. Subsidies increase when the price cap is reduced below the VOLL, or when new production units display an intermittent supply. This explains a paradoxical finding: taxing greenhouses gases may optimally be accompanied by higher subsidies to gas plants.

We have also emphasized that carbon taxation has effects not only on supply, but also on demand. The electrification of uses makes electricity even more indispensable, and blackouts even more costly. We expect that the management of blackouts will become a central topic of electricity markets.

Whether our simple and effective regulation is robust to market power remains to be studied. On the other hand, it is easily understood that all these regulatory efforts stem from the inability to regulate demand in real-time. Any progress in this direction would not only reduce the probability of rationing, and the inefficiencies associated to the price cap; it would also save some investments that are only needed due to artificially high peaks; and it would also reduce market power, by making demand more elastic. Real-time pricing is thus likely to be the relevant frontier for efficiency gains, and for alleviating the weight of regulation. In particular, innovative retail contracts could contribute usefully to this change.

A Details in the derivations in Section 2

Total expected surplus is

$$E\Big[m^s(\kappa(s))v^s(D^s(\overline{p})) + (1-m^s(\kappa(s)))v^s(0) - n^s(\kappa(s))C^s\Big(\frac{m^s(\kappa(s))}{n^s(\kappa(s))}D^s(\overline{p}), X\Big)\Big] - \sum_k I_k(x_k)$$

that we maximize wrt $(\kappa, \overline{p}, X)$, under the constraints (with multiplier $\beta(s)$ in each state)

$$\kappa(s)D^s(\overline{p}) = \sum_k x_k K_k^s = K^s(X).$$

Let us first provide necessary conditions associated to the choice of $\kappa(s)$. Note that for the sake of clarity, we often omit arguments in the formulas below. In states where all consumers are served ($m = 1$), the Lagrangian reduces to

$$v(D) - n(\kappa)C\left(\frac{D}{n(\kappa)}, X\right) - \beta\kappa D.$$

Use the identity $\frac{\partial n}{\partial \kappa} = \nu n/\kappa$ from (4), and divide by D to obtain Equation (7):

$$\text{If } m = 1, \qquad \beta(s) = \frac{\nu}{\kappa}\Big[C_Q(\frac{D}{n}) - AC(\frac{D}{n})\Big].$$

In states in which a blackout occurs ($m < 1$), the Lagrangian reduces to

$$n(\kappa)\kappa(v(D) - v(0)) - n(\kappa)C(\kappa D, X) - \beta\kappa D.$$

Use once more the identity $\frac{\partial n}{\partial \kappa} = \nu n/\kappa$, and divide by D to obtain Equation (8):

$$\text{If } m < 1, \qquad \beta(s) = n\Big[\ell(D) - C_Q(K) + \nu(\ell(D) - AC(K))\Big].$$

Regarding investments, the first-order condition with respect to the number x_k of units of type k can be simplified using Property (1) to derive Equation (9):

$$I_k'(x_k) = E\Big[n\pi_k(C_Q)\Big] + E[\beta K_k].$$

We finally turn to the first-order condition with respect to the retail price $\overline{p}$, in which we use the identities $D_p = -\varepsilon D/\overline{p}$ and $\kappa D = K$ to get Equation (10):

$$E\Big[\varepsilon m D(\overline{p} - C_Q)\Big] = E[\varepsilon\beta K].$$

This concludes this part. □

B Proof of Proposition 1

Equation (10) characterizing the optimal retail price reads

$$E\Big[\varepsilon m D(\overline{p} - C_Q)\Big] = E[\varepsilon \beta K].$$

The left-hand side can be decomposed into

$$E\Big[\varepsilon m D(\overline{p} - p)\Big] + E\Big[\varepsilon m D(p - C_Q)\Big].$$

From Properties (11) and (12) for the wholesale price p, the second term is zero when $m = 1$; and when $m < 1$, then $mD = nK$ and $p = P$, so that this second term reduces to

$$E\Big[1_{m<1} \varepsilon n K(P - C_Q(K))\Big],$$

so that we have:

$$\overline{p}E\Big[\varepsilon m D\Big] + E\Big[1_{m<1}\varepsilon n K(P - C_Q(K))\Big] = E[\varepsilon \beta K] + E\Big[p \varepsilon m D\Big].$$

Comparing to Equation (14), we obtain the value of the optimal retail tax:

$$\tau E\Big[\varepsilon m D\Big] = E\Big[\varepsilon \beta K\Big] - E\Big[1_{m<1}\varepsilon n K(P - C_Q(K))\Big].$$

The result in the Proposition follows, from (19).

We proceed similarly for the optimal subsidies. For each type k of production unit, Equation (9) characterizing the optimal investments reads:

$$I_k'(x_k) = E\Big[n \pi_k(C_Q)\Big] + E[\beta K_k].$$

Let us first analyze profits. When $m = 1$, from Property (11), we have $p(s) = C_Q$, so that

$$\pi_k(C_Q) = \pi_k(p(s)).$$

When $m < 1$, we have $p(s) = P$, and the unit produces at full capacity both when the price is P and when the price is $C_Q(K)$. Therefore,

$$\pi_k(C_Q(K)) = C_Q(K)K_k - c_k(K_k) = (C_Q(K) - P)K_k + \pi_k(p(s)).$$

Replacing, we obtain

$$I_k'(x_k) = E\Big[n\pi_k(p(s))\Big] + E[\beta K_k] - E\Big[1_{m<1}n(P - C_Q(K))K_k\Big].$$

Compare to the Equation (15) characterizing the equilibrium investments to get

$$\sigma_k = E[\beta K_k] - E\Big[1_{m<1}n(P - C_Q(K))K_k\Big] = E[\Delta K_k].$$

This concludes the proof. □

C A study of the case when demand is impacted by a multiplicative shock

Recall the definition of this case in Definition (20):

$$D(p, s) = b(s)d(p),$$

where $b > 0$ and the function d is decreasing. This demand function derives from the following surplus function

$$v(e, s) = a(s) + b(s)G\left(\frac{e}{b(s)}\right),$$

where G is a primitive of the inverse function d^{-1} of d. Now, it is easily seen that the elasticity of demand

$$\varepsilon^s(p) = \frac{-pd'(p)}{d(p)}$$

and the VOLL measured at the consumption point

$$\ell^s(D^s(p)) = \frac{G(d(p)) - G(0)}{d(p)}$$

are both independent of s, as announced. Moreover, ε simplifies in Equation (14), so that Constraint (13) implies that the competitive access charge is exactly zero, as announced. Finally, ε also simplifies in Equation (17), which reduces to

$$\tau E[mD] = E[\Delta K],$$

and the right-hand side is exactly $\sum_k \sigma_k x_k$, from Equation (18). This proves that the regulatory budget is balanced, as announced.

D Numerical simulations

We present here the fictitious economy whose functioning is simulated numerically. As mentioned in the text, the representative consumer displays a surplus with a constant elasticity of substitution:

$$s^D S\left(\frac{e}{s^D}, \frac{g}{s^D}\right) = a s^D \left(b\left(\frac{e}{s^D}\right)^{\rho} + c\left(\frac{g}{s^D}\right)^{\rho}\right)^{\frac{\gamma}{\rho}}$$

with $a = 8.1$, $b = 3$, $c = 1$, $\rho = 0.65$, $\gamma = 0.52$. These values ensure that S is concave, and that the two goods are substitute in each state of nature. The demand shock s^D is uniform on the interval $[0.3, 0.8]$.

Nominal capacities are $K_1 = K_2 = 1$, but the renewable capacity K_1 is affected by a multiplicative shock s^S, assumed uniform on $[0.3, 1]$. Investment costs are quadratic for renewables and linear for gas: $I_1(x_1) = x_1^2$ and $I_2(x_2) = 1.5x_2$. In this framework with two energy sources, the aggregate cost function in state s^S equals

$$C^{s^S}(Q, X) = \begin{cases} 0 & \text{for } Q \leq s^S x_1 K_1 \\ c_2\left(Q - s^S x_1 K_1\right) & \text{for } s^S x_1 K_1 \leq Q \leq s^S x_1 K_1 + x_2 K_2 \\ c_2 x_2 & \text{for } Q \geq s^S x_1 K_1 + x_2 K_2. \end{cases}$$

The profit function of each unit equals

$$\pi_1^{s^S}(p) = p s^S x_1 K_1 \text{ and } \pi_2(p) = \max(p - c_2, 0) x_2 K_2.$$

The proportion $n^s(\kappa)$ of connected production units is (with $\overline{\kappa}$ set at 0.95):

$$n(\kappa) = \begin{cases} \frac{\kappa}{\overline{\kappa}^2}\left(2\overline{\kappa} - \kappa\right) & \text{for } \kappa \leq \overline{\kappa} \\ 1 & \text{for } \kappa > \overline{\kappa} \end{cases}$$

References

Abbott, Malcolm. 2001. "Is the Security of Electricity Supply a Public Good?" *The Electricity Journal*, 14(7): 31–33.

Ambec, Stefan, and Claude Crampes. 2019. "Decarbonizing Electricity Generation with Intermittent Sources of Energy." *Journal of the Association of Environmental and Resource Economists*, 6(6): 1105–1134.

Borenstein, Severin, and Stephen Holland. 2005. "On the Efficiency of Competitive Electricity Markets with Time-Invariant Retail Prices." *RAND Journal of Economics*, 36(4): 469–493.

Borenstein, Severin, James Bushnell, and Erin Mansur. 2023. "The economics of electricity reliability." *Journal of Economic Perspectives*, 37(4): 181–206.

Butters, R .Andrew, Jackson Dorsey, and Gautam Gowrisankaran. 2024. "Soaking up the sun: Battery investment, renewable energy, and market equilibrium." National Bureau of Economic Research.

Elliott, Jonathan T. 2024. "Investment, Emissions, and Reliability in Electricity Markets." *wp.*

Fabra, Natalia. 2018. "A primer on capacity mechanisms." *Energy Economics*, 75: 323–335.

Ferrasse, Jean-Henry, Nandeeta Neerunjun, and Hubert Stahn. 2022. "Intermittency and electricity retailing: An incomplete market approach." *Mathematical Social Sciences*, 120: 24–36.

Gerlagh, Reyer, Matti Liski, and Iivo Vehviläinen. 2024. "Rational Rationing: A Price-Control Mechanism for a Persistent Supply Shock." *wp.*

Gorman, Will. 2022. "The quest to quantify the value of lost load: A critical review of the economics of power outages." *The Electricity Journal*, 35(8): 107–187.

Gowrisankaran, Gautam, Ashley Langer, and Wendan Zhang. 2025. "Policy Uncertainty in the Market for Coal Electricity: The Case of Air Toxics Standards." *Journal of Political Economy*, 133: 1757–95.

Gowrisankaran, Gautam, Stanley S. Reynolds, and Mario Samano. 2016. "Intermittency and the value of renewable energy." *Journal of Political Economy*, 124(4): 1187–1234.

Haar, Lawrence. 2021. "The competitive disadvantages facing British assetless electricity retailers." *Energy Policy*, 155: 112323.

Holland, Stephen P., Erin T. Mansur, and Andrew Yates. 2022. "Decarbonization and Electrification in the Long Run." *NBER Working Paper 30082*.

Joskow, Paul, and Jean Tirole. 2006. "Retail electricity competition." *The RAND Journal of Economics*, 37(4): 799–815.

Joskow, Paul, and Jean Tirole. 2007. "Reliability and competitive electricity markets." *The RAND Journal of Economics*, 38(1): 60–84.

Joskow, Paul L. 2022. "From hierarchies to markets and partially back again in electricity: responding to decarbonization and security of supply goals." *Journal of Institutional Economics*, 18(2): 313–329.

Keppler, Jan Horst. 2017. "Rationales for capacity remuneration mechanisms: Security of supply externalities and asymmetric investment incentives." *Energy Policy*, 105: 562–570.

Léautier, Thomas-Olivier. 2016. "The visible hand: ensuring optimal investment in electric power generation." *The Energy Journal*, 37(2): 89–109.

Leslie, Gordon, Armin Pourkhanali, and Guillaume Roger. 2021. "Can real-time pricing be progressive? Identifying cross-subsidies under fixed-rate electricity tariffs." *Identifying Cross-Subsidies under Fixed-Rate Electricity Tariffs (January 27, 2021)*.

Llobet, Gerard, and Jorge Padilla. 2018. "Conventional Power Plants in Liberalized Electricity Marketswith Renewable Entry." *The Energy Journal*, 39(3): 69–92.

McRae, Shaun D, and Frank A Wolak. 2019. "Market power and incentive-based capacity payment mechanisms." *Unpublished manuscript, Stanford University.*

Wolak, Frank A. 2013. "Economic and political constraints on the demand-side of electricity industry re-structuring processes." *Review of Economics and Institutions*, 4(1): 42.

Wolak, Frank A. 2022. "Long-term resource adequacy in wholesale electricity markets with significant intermittent renewables." *Environmental and Energy Policy and the Economy*, 3(1): 155–220.