ifo WORKING PAPERS

342
2020
November 2020
(Revised September 2021)

Efficient Pricing of Electricity Revisited

Mathias Mier

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

Imprint:

ifo Working Papers
Publisher and distributor: ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49(0)89 9224 0, Telefax +49(0)89 985369, email ifo@ifo.de
www.ifo.de

An electronic version of the paper may be downloaded from the ifo website:
www.ifo.de

ifo Working Paper No. 342

Efficient Pricing of Electricity Revisited

Abstract

Increasing shares of intermittent renewable energies challenge the dominant way to trade electricity ex-ante in forward, day-ahead, and intraday markets: Coal power plants and consumers cannot react to the stochastic element of renewables, whereas gas turbines can. We use a theoretical model to analyze behavior of final consumers and incentives of perfectly competitive firms to invest in different types of technologies under ex-ante pricing. Curtailed consumers need to get compensated in high of their disruption cost. Coal power firms recover cost. Renewables and gas turbine firms fail. We identify imperfections that arise from the delay in price setting and market clearing. Do real-time prices induce an efficient outcome? Consumers need to get taxed in high of rationing cost. Support is redundant for gas turbine firms, but renewables firms still fail to recover cost because the spatially distributed nature of renewables creates an output risk.

JEL Code: D41, D47, Q41, Q48, L94, L98
Keywords: Efficient pricing, market design, capacity mechanisms, renewable energies, supply uncertainty, consumer behavior

Mathias Mier
ifo Institute – Leibniz Institute for
Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
Phone: +49-89-9224-1365
mier@ifo.de

1 Introduction

A resolute decarbonization of the economies around the world is crucial to limit the probability of global warming above 2°C to significantly reduce the risks and impacts of climate change (Paris Agreement, 2015). Countries started with decarbonizing the power sector, because renewable energies easily substitute for carbon-intensive technologies, and electrification allows one to decarbonize other sectors (e.g., transport, heat, heavy industry via green hydrogen). This article studies consumer behavior and investment incentives of different types of technologies (e.g., coal power, wind power, and gas turbines) under *ex-ante* (e.g., over-the-counter, forward, day-ahead, intraday) and *real-time pricing* detect and resolve market imperfections that arise from the delay in price setting (ex-ante) and market clearing (real time), as well as from the spatially distributed and stochastic nature of renewable energies. The social planner solution serves as a benchmark to analyze the consumption of non-reactive consumers and investment behavior of perfectly competitive firms. We derive compensation payments for final consumers when consumption is inefficient and technology-specific support mechanisms to incentivize the efficient outcome when firms fail to recover cost or do not install first-best capacities. We finally apply our theoretical findings on real-world power market designs and reveal remaining flaws that prevent efficient pricing of electricity.

Coal and nuclear power plants must be scheduled ahead of actual production because ramping up and down takes a long time or is too costly. Intermittent renewables randomly fluctuate at zero marginal cost, and require gas turbines to balance these fluctuations due to the lack of demand response in electricity markets.[1] The key challenge in transitioning from a fossil-fuel dominated to a renewables dominated power system is to ensure *resource adequacy*, meaning that installed capacities are sufficient to supply the electricity demanded by consumers at all times—which is a private good.[2] In fossil-fuel dominated power systems, forward markets or over-the-counter trading ensure generators against demand risks, day-ahead markets address the limited dispatchability of nuclear and coal power plants, intraday trading accounts for the declining level of demand uncertainty, and balancing (in Europe) or real-time (in Northern America) markets, respectively, settle the remaining deviation between supply and demand. Additionally, reserve markets or markets for ancillary services ensure security of services, meaning the physical integrity of the system—which is a public good.

[1] Most electricity markets face low demand flexibility (Joskow and Tirole, 2006, 2007), in particular, in real time (Joskow, 2011). Bejan et al. (2019) found that consumers suffer utility losses from reacting to real-time prices. There is another factor that drives inelasticities: retail prices and wholesale prices are hugely disconnected because the price that consumers face is largely based on taxes, levies, and network fees. Moreover, retail prices are additionally distorted by the fact that they do not reflect fluctuating demand (and supply) within a day (Borenstein and Bushnell, 2018).

[2] Other commodity markets delay the timing between production and delivery by using (producer and consumer) storage, but electricity storage is yet economically not beneficial (e.g., batteries or power-to-gas) or scarce regarding suitable sites (e.g., pumped hydro).

Higher shares of renewable energies lead to extended periods of *zero prices* and narrow the business case of coal and nuclear power due to *ramping constraints* and related cost. The final uncertainty about the amount of wind and solar power in the system realizes in real time, which increases the demand for *generator flexibility* and the related price uncertainty. These factors diminish the capability of dispatchable generators to recover cost, and challenge whether or not current market design can deal with fluctuating renewables and provide resource adequacy (e.g., Newbery et al., 2018; Joskow, 2019). Additionally, the settlement volume of reserve markets or markets for ancillary services increased as a consequence of rising shares of intermittent renewables, leading to an implicit subsidization of fluctuating renewables or inflexible generators, respectively. In response, countries started to implement additional capacity mechanisms, but whether or not this is the first-best solution remains unclear.

Our model introduces three types of technologies to reflect that supply of *renewables* (e.g., wind and solar) is spatially distributed and stochastic, *steam power* (e.g., nuclear and coal power plants) cannot react to the stochastic fluctuations of renewables, whereas *gas power* (e.g., gas turbines and combined-cycle power plants) can. Demand is steady due to ex-ante fixed consumer prices or the lack of real-time demand response, respectively. Long-term capacity decisions are made years ahead of actual delivery. Demand is fixed at least some hours ahead of actual delivery. Production of steam power must be scheduled some hours ahead of actual production as well. The supply of renewable energies realizes in real time, and renewables and gas power are dispatched. Finally, consumers are curtailed and suffer disruption cost when production is not sufficient to meet demand.[3]

Our theoretical results are threefold. First, under ex-ante pricing, steam power firms recover cost, whereas gas power firms fail. We identify inadequately priced balancing as the inefficiency caused by the temporal deviation between price setting and market clearing. Steam power does not contribute to balance stochastic variations in renewable's supply due to *ramping constraints*. Gas power adjusts production according to renewables feed-in, that is, gas power faces an output risk that is not reflected in the price. A combination of capacity payments and payments for the energy supplied (so called flat feed-in tariffs) addresses *generator flexibility* of gas power. Payments that are conditional on the realization of wind and solar availability (so called state-contingent feed-in tariffs) even avoid capacity payments.

Second, real-time pricing instantly resolves the problem of pricing gas power's ability to adjust production. Gas power still faces an output risk but this risk is adequately reflected in real-time prices that rise above the value of lost load (VOLL) in lost load events. In contrast, renewables

[3] See Küfeoğlu et al. (2015) for an evaluation of consumer interruption (disruption) cost (in Finland) and Oren and Doucet (1990) for an insurance scheme when consumers are differentiated by their disruption cost. Moreover, if demand would be sufficiently responsive, the market would always clear, and consumers would never suffer involuntary curtailment (Cramton et al., 2013).

recover cost and provide first-best capacities when and only when random output from a generating unit is perfectly correlated to total renewables output. When outputs are not perfectly correlated or even independent, markets are incomplete because firms cannot ensure against their output risk. There is no direct spatiality in the theoretical model, but we can interpret a perfect correlation as follows: Each wind turbine and each solar PV panel in the system faces the same weather conditions no matter what location. This assumption is realistic only for spatially small market areas with homogeneous weather conditions. Under such a setup, renewables firms do not face an output risk because every renewable firm faces the same weather conditions (perfect correlation). For spatially bigger markets, the assumption of perfectly correlated generating units might turn realistic when flexibly clustering zones that face the same weather conditions and apply zonal pricing. Moreover, there is no inefficiency on the balancing side of the market (that is, for flexible generators). Additional capacity mechanisms or the usage of reserve or ancillary service markets (against their original intention) to improve resource adequacy are superfluous under real-time pricing.

Third, turning to behavior of final electricity consumers such as private households, companies, or industrial consumers, curtailed consumers need compensation in high of disruption cost when paying ex-ante prices.[4] Similar as gas power does, consumers offer system flexibility and need compensation. When consumers pay real-time prices, efficient consumption demands for taxing curtailed consumers in high of rationing cost.[5] In contrast to gas power, consumers do not reduce the overall burden of rationing cost. Additionally, efficient prices rise (by rationing cost) above the value of lost load (VOLL) of the last curtailed consumer in lost load events. Consumers with a VOLL below the resulting price and above the VOLL of the last curtailed consumer would be better of in reducing their demand below the efficient level. The tax thus imposes opportunity cost of not demanding electricity.

Finally, we apply our theoretical findings to analyze potential flaws in real-world market designs. Under the assumptions of our model, the combination of day-ahead and real-time markets (as in Northern America) is equivalent to a combination of ex-ante and energy-only balancing markets, and lead to the efficient solution (when renewables generating units are perfectly correlated). Even balancing markets that provide payments for capacity (as in Europe) can establish the efficient outcome. However, in all those markets pricing is efficient if and only if generating firms pay undersupply penalties in high of consumers' disruption cost for each unit of undersupply and

[4] We are aware that final consumers do not interact with firms on day-ahead or real-time markets, respectively. Intermediary retailers close that gap by buying from firms on wholesale markets and selling at respective (i.e., ex-ante or real-time) prices to final consumers. We analyze such a constellation in Subsection 6.3 and find that the market outcome is equivalent to one where final consumers would directly purchase electricity on wholesale markets or generating firms, respectively. For parsimony, we thus neglect the role of intermediaries for most parts of the article.

[5] Rationing cost are cost of the system operator to curtail those consumers with the lowest marginal utility

carry the burden of rationing cost (when final consumers pay ex-ante prices) or do not pay for undersupply nor rationing cost (when final consumers pay real-time prices). Moreover, final consumers require compensation in high of disruption cost in case of lost load (when final consumers pay ex-ante prices) or get taxed by rationing cost (when final consumers pay real-time prices).

Section 2 provides a review of the related literature. We then introduce the basic model in Section 3. Section 4 describes the social planner solution, which serves as a benchmark for the following analysis of efficient pricing. Section 5 shows how cost recovery can be achieved under ex-ante pricing by support mechanisms, analyzes cost recovery under real-time pricing, and shows the behavior of final consumers. Section 6 analyzes real-world market designs and suggests how to tackle remaining flaws. Section 7 concludes by giving guidance for future electricity market design.

2 Related Literature

Recent studies discuss the need of capacity mechanisms due to the increasing shares of randomly fluctuating renewables and the problems of dispatchable generators such as gas and steam power to keep profitable (e.g., Cramton, 2017; Wolak et al., 2020). The core idea of capacity mechanisms—incentivizing capacity investments to obtain efficient system reliability and consumer rationing—is born from the peak-load pricing literature.[6] The innovation was also to charge consumers based on capacity cost. Deterministic models provide sufficient rules for *efficient pricing* (e.g., Houthakker, 1951; Hirshleifer, 1958; Williamson, 1966; Turvey, 1968). The literature tends to focus on different *rationing schemes* as soon as demand uncertainty enters the picture. Brown and Johnson (1969) assume that consumers are served regarding their willingness-to-pay (perfect load shedding), whereas Crew and Kleindorfer (1976) assume that perfect load shedding causes rationing cost. Visscher (1973) describes two alternative approaches. Either consumers are served randomly, or consumers with the lowest willingness-to-pay (WTP) are served first. Turvey and Anderson (1977) implement constant marginal cost of lost load (also *excess demand* in the economic literature) and abstract from surplus losses due to lost load. Chao (1983) chooses the same setup, but he additionally assumes that generating units are subject to random failures that are stochastically independent of one another. Kleindorfer and Fernando (1993) model supply uncertainty in the same way, but they additionally account for surplus losses from lost load and distinguish between rationing and disruption cost.[7] As we do, Eisenack and Mier (2019) choose the same specification of supply uncertainty by additionally considering the case of perfectly correlated generating units. They opt

[6] The peak-load pricing literature was developed by Bye (1926, 1929); Boiteux (1949); Steiner (1957). Crew et al. (1995) provide an excellent survey.

[7] Rationing cost are dedicated to the system operator to obtain perfect load shedding, and disruption cost are direct cost to consumers.

for Chao's rationing approach, whereas we use the Kleindorfer and Fernando formulation.

Whether these efficient pricing rules lead to cost recovery has not been much studied to date. In all deterministic setups, firms recover cost, but for the stochastic models, the outcome is diverse. Perfect load shedding with zero rationing cost (Brown and Johnson, 1969), as well as random rationing with additive demand uncertainty (Visscher, 1973), lead to prices below long-run marginal cost (LRMC)—so that cost recovery is not possible. Carlton (1977) shows that random rationing with multiplicative demand uncertainty would permit cost recovery. Serving consumers with the lowest WTP first would even lead to prices above LRMC and strictly positive profits are possible. Chao (1983) distinguishes between two polar cases: marginal demand that is independent from total demand and marginal demand that is perfectly correlated to demand. Prices are too low to recover cost for the independence case. Results are inconclusive for the correlation case. The most comprehensive analysis is by Kleindorfer and Fernando (1993). Additive demand uncertainty will lead to a price weakly below LRMC, whereas multiplicative demand uncertainty might lead to a price above LRMC. It is difficult to conclude whether or not cost recovery is possible.

How to actually implement these rules, for example, by market design and support mechanisms, is not touched upon in the peak-load pricing literature. The standard theory suggests that varying prices for electricity on the basis of short-run marginal cost (SRMC) with a price cap at the average value of lost load (VOLL) provide sufficient incentives for capacity investments (e.g., Bushnell et al., 2017).[8] However, power systems are subject to many possible market failures. Distribution and transmission of electricity are natural monopolies, so grids are regulated in most power systems, whereas power generation and retail supply are organized in an energy-only market. Joskow and Tirole (2007) point to the problem of pricing security of service, the ability of the system to withstand sudden disturbances (Antweiler, 2017). Security of service is a public good and private provision would lead to undersupply. System operators use markets for ancillary services to ensure security of service. These cost are often not internalized in the final price for consumers but are contained in network fees, which are often averaged across consumers.[9] Thus, consumers with higher preferences for security are free-riding on the consumers with lower ones. This is related to imperfections that might arise from asymmetric information, for example, about the preference for security of service and the VOLL. Risk-averse policy makers tend to overestimate the VOLL and provide inefficiently high back-up capacity (Crampes and Salant, 2018); and risk-averse investors underinvest due to incomplete markets (e.g., Neuhoff and De Vries, 2004). To prevent the abuse of market power in scarcity events, price caps below the average VOLL

[8] The VOLL are the social cost of electricity shortages and entail surplus losses due to unconsumed electricity as well as disruption cost, reflecting the willingness-to-accept curtailment.

[9] We understand ancillary services as the quality of service at a particular location to meet physical network requirements. Note that nodal pricing of electricity would allow the market to internalize at least cost from transmission and related ancillary services.

lead to missing money.[10] Finally, environmental externalities should be addressed by a Pigouvian tax, and positive externalities from induced technological change require subsidization.

The author is not aware of a theoretical model that covers all these market failures. Existing studies always focus on one or two of them. We use a reference model that abstracts from the imperfections discussed and focuses on the general issue that an energy-only market (as in our model) could lead to missing money and under-procurement of capacity, for example, due to insufficiently priced balancing and flexibility (Newbery, 2016). Thus, the article contributes to the discussion on whether or not capacity mechanisms—as lately implemented by, for example, Great Britain, France, and Australia—are necessary (e.g., Keppler, 2017; Milstein and Tishler, 2019; Bublitz et al., 2019).

Similar to the focus of the article, Chao (2011) derives efficient pricing and capacity rules under ex-ante and real-time pricing with a dynamic demand response. He considers a renewables technology, whose (uncertain) supply is inversely correlated with demand, which is a key difference from our article regarding theoretical modeling.[11] He considers surplus losses and rationing/disruption cost in the theoretical part of his work, but abstracts from them to derive results using a numerical simulation with wind power, gas turbines, and combined-cycle power plants. Chao's model is more comprehensive than ours with regard to demand modeling; our model captures the fact that technologies differ in dispatchability. Moreover, he has to rely on the numerical simulation to obtain interpretable results, whereas we can derive the necessary design of support mechanisms and markets directly from the theoretical model.

One of our core findings is that, under real-time pricing, support mechanisms are needed only for renewable energies (and consumers). Most literature had focused on a combination of feed-in tariffs and capacity payments for renewable energies (e.g., Lesser and Su, 2008). Model-wise, most closely related to that skein of literature is Antweiler (2017). He derives optimal pricing instruments by taking into account the correlation between renewables sources and between these sources and demand. The main difference from our approach is that we account for the correlation in the availability of single solar PV panels or wind turbines due to weather conditions. We can confirm his finding that capacity payments or flat feed-in tariffs alone do not provide optimal investment incentives, but additionally suggest state-contingent feed-in tariffs that would make combined instruments for energy and capacity obsolete.

Other related articles are Ambec and Crampes (2012) and Helm and Mier (2019). Ambec and

[10] A price cap suppresses prices and thus leads to excess demand by consumers (Leautier, 2018; Helm and Mier, 2019). See Bulow and Klemperer (2012) for a broader view on price caps, Joskow and Tirole (2007) for positive effects and Hogan (2005) for negative effects, as well as Fabra et al. (2011); Zöttl (2011); Schwenen (2014) for electricity market applications. More recently, Fabra (2018) suggests a price cap to avoid the development of market power and a capacity payment as a contrary instrument. Bajo-Buenestado (2017) finds that price caps lead to an inefficient market outcome in a perfectly competitive market, which could turn under market power.

[11] See Pinho et al. (2018) for another stylistic model with an uncertain supply of renewable energies.

Crampes (2012) provide a model with a fossil and a renewables technology, whose deterministic availability is either 0 or 1. An ex-ante price for non-reactive consumers leads to overinvestment and no profit in the fossil technology, but underinvestment and profit in the renewables technology, whereas the technology we investigate that is most similar to their fossil one cannot recover cost. The issue of cost recovery could be solved by structural integration of different technologies within a single company. They do not contribute to the debate over capacity mechanisms because, in contrast to our article, they assume that reliable capacity must be sufficient to meet demand. The same assumption is made by Helm and Mier (2019), who allow the renewables technology to take any value between a minimum value and 1. They combine reactive consumers that are subject to dynamic pricing and non-reactive consumers that face an ex-ante price. In their model, the efficient solution could be decentralized. In particular, they identified a capacity premium for fossil generators provided through the market. However, their focus is on the efficient diffusion pattern of renewable energies and policies are considered only in the case of a price cap. because dynamic pricing is possible in their setup, they abstract from the core issues of this article: how to avoid capacity mechanisms by market design and pricing schemes.

None of the articles just discussed, with the exception of Eisenack and Mier (2019), account for different levels of dispatchability. The topic is only addressed by engineering dispatch (e.g., Kumano, 2011) and empirical models (e.g., Novan, 2015), although it is highly relevant for the integration of large shares of renewables (see Schill et al., 2017 for start-up cost of thermal power plants). Perhaps the theoretical model that comes closest to this topic is that of Green and Léautier (2017) who study inflexibilities—modeled by minimum production levels—of renewables and of conventional capacity. In a numerical simulation they find that higher wind turbine capacities squeeze out the inflexible more quickly than do flexible nuclear generators. Another theoretical model by Crampes and Renault (2019) uses ramping rates to model the adaptation speed of technologies.

Finally, our article is related to the literature about undersupply penalties (e.g., Sunar and Birge, 2019) when firms are not able to deliver their ex-ante pledge and consumer compensation schemes in lost load events. Undersupply penalties are largely implemented but schemes differ hugely in the penalty level.[12] We suggest that penalties should be equal to (efficient) market prices, which raise above the VOLL (by rationing cost) in lost load events. Compensation schemes for final consumers in turn are only rarely implemented and not related to potential undersupply of generating firms but rather due to management and investment flaws on the distribution level. For example, Sweden (Wallnerström, 2008) and United Kingdom have institutionalized compensations schemes for durable lost load events, and several Northern American utilities pay compensations as well (Costello, 2012). However, we suggest that compensations are paid not only for durable outages or management flaws on the distributional level, respectively, but for each unit of lost load

[12] EC (2016) gives on overview how European markets deal with undersupply or imbalance settlements, respectively.

caused by undersupply of generating firms. The reasoning is the same: provision of the efficient amount of reliability or resource adequacy, respectively.

3 Model

Consider three types of technologies: $j = r, s, g$. r are renewables technologies like wind turbines and solar PV with random supply; s are steam power technologies, which are limitedly dispatchable. Scheduling steam power requires planning a certain time ahead of actual production; g are gas power technologies, which are perfectly dispatchable because they can adjust production instantly.

We consider an investment cycle with one period of production and consumption. We could easily extend the one-period model by considering multiple periods of production and consumption, but the notational overhead comes at enormous expense for the same insight. However, we perform that task in Section 5.2 to show the feasibility and possible generalization of the model.

Capacity is Q_j, production is Y_j, and $\tilde{Q}_r$ is the random available capacity of renewables. It is convenient to denote aggregate production by $Y = \sum_j Y_j$. Production is restricted by (available) capacity, i.e., $Y_s \leq Q_s, Y_g \leq Q_g, Y_r \leq \tilde{Q}_r$. Demand is D, and $L = \max\{D - Y, 0\}$ is *lost load* (or excess demand), so that $X = \min\{D, Y\}$ is final consumption. Consumers obtain utility U from consumption. Utility is concave, i.e., $U' > 0$, $U'' < 0$, and fulfills Inada conditions.

In case of lost load, consumers suffer *utility losses* and additionally *disruption cost*.[13] The system operator curtails the consumption with the lowest marginal utility at the burden of *rationing cost*.[14] Thus, the utility loss from lost load is given by the difference between utility from scheduled demand and utility from final consumption, i.e., $U(D) - U(X) \geq 0$. We assume that marginal disruption cost c_L and marginal rationing cost δ are constant. The underlying assumption is that all (curtailed) consumers are homogeneous with regard to disruption cost and the effort to curtail those with the lowest marginal utility first, otherwise the assumption of constant c_L, δ is not credible.[15]

Wind turbines and solar PV panels are the most promising renewable energy source to decarbonize economies. Their joint features are a small unit size, random availability, and spatial distribution. We thus construct available capacity of renewables as a boundedly integrable random variable and define it as $\tilde{Q}_r := \int_0^{Q_r} \omega(z)\, dz$; i.e., $\tilde{Q}_r$ is conceived as a continuum of marginal generating units z with random availabilities $\omega(z) \in [0, 1]$. Generation units z are stochastically identically distributed random variables, meaning that all marginal generating units face the same climate. However, actual weather conditions might differ due to the spatial distribution of, e.g., wind turbines and

[13] For example, in food industries lost utility reflects the value of food that cannot be produced due to lost load. However, electricity shortages might even destroy food that is already produced due to a disruption of the cooling chain.

[14] See Crew and Kleindorfer (1976) for the seminal contribution and Kleindorfer and Fernando (1993) for a discussion of rationing, disruption, and utility losses by curtailment.

[15] Consumers could still be heterogeneous with regard to their obtained utility.

solar PV. The correlation between those units covers varying weather conditions due to spatial distribution. We consider two extreme cases. The *independence case* is when the availability of a (marginal) generating unit $\omega(z)$ realizes stochastically independently from total available capacity $\tilde{Q}_r$, which we denote *ind*.[16] In the *perfect correlation case*, denoted *corr*, the availabilities of marginal generating units are perfectly correlated to $\tilde{Q}_r$. The intuition is as follows: When weather conditions are always the same for each generating unit, we are in the case of perfect correlation. When weather conditions are independent from each other, we are in the case of independence. Of course, these cases are extremes and the reality is inbetween. However, considering these two extreme cases allows us to derive results applicable to more realistic settings that are neither *ind* nor *corr*.[17]

Short-run marginal production cost (SRMC) c_j and marginal capacity cost b_j are constant. Renewables have the lowest or even zero SRMC, whereas gas power has the highest, i.e., $0 \leq c_r < c_s < c_g$. Pollution externalities for steam and gas power are internalized within SRMC, for example, by means of Pigouvian taxation.[18] We further assume that the long-run marginal cost (LRMC) of gas power are the highest. Otherwise, it might be beneficial to use gas power only. Producing one unit with gas power is cheaper than accepting one unit of lost load, i.e., $b_g + c_g < c_L$. Otherwise, lost load dominates gas power production. Nevertheless, the choice between accepting lost load and providing (often unused) back-up capacities is the outcome from maximizing the difference between surplus from consumption, $U(X) - (c_L + \delta)L$, and cost, $\sum_j \left(b_j Q_j + c_j Y_j\right)$.

The timing of decisions is shown in Figure 1. The essential assumption is that capacities are fixed in the short-run, and demand must be decided before production decisions are made under technological restrictions. Thus, in Stage 1, a social planner selects capacities $Q_j, j = r, s, g$ and, in Stage 2, demand D via setting a price. Steam power production $Y_s \leq Q_s$ needs to be specified before the random availability of renewables $\tilde{Q}_r$ realizes and cannot be changed later (Stage 3). This places steam power earliest in dispatch timing. To assure that this placement is credible, we assume that the cost of ramping down steam power is higher than the benefits of replacing steam power with renewables production.[19] Similarly, ramp-ups are technologically not possible in the

[16] Note that we do not account for the correlation of different renewables sources nor the correlation of renewables availability with demand (there is no demand uncertainty in the model). We solely concentrate on the correlation of single wind turbines (or solar PV panel clusters) with total wind turbine (or solar PV) availability. Geographically small power system (or pricing zones) would have correlations close to the perfect correlation case, whereas bigger power systems (or pricing zones) might tend to the independence case.

[17] It is possible to implement a correlation measure as done in Chao (1983) for the correlation of marginal demand with total demand. The outcome would be a weighted average of correlated and uncorrelated results. Thus, the additional insight for this analysis from such a correlation measure is limited without knowing real-world correlation and conducting numerical simulations.

[18] See Helm and Mier (2021) for an analysis of optimal renewables subsidies when Pigouvian taxation is imperfect.

[19] This mechanism reflects the often observed behavior of steam power plants to bid negative prices to avoid ramping down (due to ramping cost).

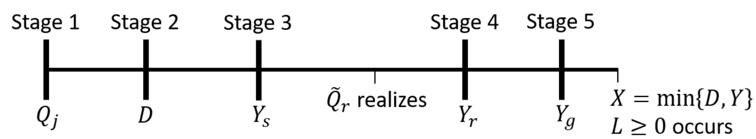

Fig. 1: Multi-stage decision structure

necessary time frame. After demand and steam power production is fixed, renewables availability realizes, and in Stage 4, production of renewables $Y_r \leq \tilde{Q}_r$ must be decided. If the actual availability of renewables is insufficient to meet demand, gas power must be employed (Stage 5), or lost load $L \geq 0$ occurs and leads to a cost when it is strictly positive. Renewables are curtailed if the renewables availability is above remaining demand.

4 Social Planner Solution

Economic theory suggests to set prices equal to marginal utility. We are going to follow this principle in Sections 5.1 and 5.2 where we determine the market equilibrium, but need to determine efficient production Y_j^*, demand D^*, and capacities Q_j^* first (asterisks denote efficient outcomes). The efficient outcome serves as benchmark to evaluate the market equilibrium, and follows from maximizing welfare,

$$W = U(X) - (c_L + \delta) L - \sum_j \left(b_j Q_j + c_j Y_j \right), \tag{1}$$

within the constraints of the decision structure shown in Figure 1. We consider only situations where $Q_r^* > 0$ to focus on the most interesting cases.[20]

Start with some preparatory notation. Assume that $a = E\left[\tilde{Q}_r\right]/Q_r \in (0,1)$ is the average availability of renewables capacity, where E is the expectation operator. $E\left[\tilde{Q}_r\right] = aQ_r$ is expected available capacity of renewables. $\Omega = [0, Q_r]$ is the sample space of $\tilde{Q}_r$ and $f\left(\tilde{Q}_r, Q_r\right)$ its probability density function. For any interval of events $\Omega_c \subseteq \Omega$, the events $\tilde{Q}_r \in \Omega_c$ realize with probability $\Pr_c := \int_{\Omega_c} f(x, Q_r)\, dx$. Call $a_c = E\left[\tilde{Q}_r | \Omega_c\right]/Q_r$ the conditional availability, where $E\left[\tilde{Q}_r | \Omega_c\right]$ is expected availability of renewables capacity on the condition that Ω_c realizes. Here, $E\left[\tilde{Q}_r | \Omega_c\right]$ is used as a shortcut to indicate $E\left[\tilde{Q}_r | \tilde{Q}_r \in \Omega_c\right]$. To avoid having to show each equation for both

[20] If $Q_r = 0$, then LRMC of steam power are lower than those of gas power. Given the fixed demand assumption, this would lead to an outcome of steam power only; at least in the one period model described here.

extreme cases *ind* and *corr*, we use the shorthand notation

$$\overline{a}_c \quad := \quad \begin{cases} a & \text{for } ind, \\ a_c & \text{for } corr. \end{cases} \tag{2}$$

Production. We now derive production decisions from Stages 3 to 5. Steam power decides production before the random availability of renewables realizes. Production should be increased under the given constraints so that $Y_s = Q_s$, as excess capacity of steam power carries no benefit in later stages.[21] Renewables have the lowest SRMC. Thus, renewables should meet as much remaining demand, $D - Q_s$, as possible after steam power production is fixed and the realization $\tilde{Q}_r$ is known. Renewables are not fully needed as long as the remaining demand is below the availability of renewables capacity. However, as soon as the available capacity of renewables is no longer sufficient to meet the remaining demand, gas power should be used to avoid lost load because producing with gas power is cheaper than accepting lost load. Gas power is dispatched when the total gas power capacity is not needed to meet remaining demand, which is $D - Q_s - \tilde{Q}_r$. The total capacity of gas power would be employed if and only if production of all technologies is not sufficient to satisfy demand.

Depending on the realization of the random variable $\tilde{Q}_r$, we distinguish between three intervals of events. *Renewables curtailed* occurs for all $\tilde{Q}_r \in \Omega_1 = [D - Q_s, Q_r]$.[22] *Gas power dispatched* realizes for all $\tilde{Q}_r \in \Omega_2 = [D - Q_s - Q_g, D - Q_s)$, and *lost load* when $\tilde{Q}_r \in \Omega_3 = [0, D - Q_s - Q_g)$. The union of the two intervals Ω_1 and Ω_2 is given by $\Omega_{12} = [D - Q_s - Q_g, Q_r]$ and realizes with probability Pr_{12}. Figure 2 illustrates the interval of events by using a stylistic probability density function of the random variable. The area under the density function is the probability that a certain interval realizes and can be calculated by using the table below the graph.

Using the definition of the intervals of events Ω_c, we can summarize efficient production in Lemma 1. Appendix A contains a comprehensive proof.

Lemma 1. *Suppose that $Q_r^* > 0$. Efficient production is given by $Y_s^* = Q_s$,*

$$\begin{aligned} E\left[Y_r^*\right] &= E\left[\tilde{Q}_r|\Omega_3\right]\text{Pr}_3 + E\left[\tilde{Q}_r|\Omega_2\right]\text{Pr}_2 + (D - Q_s)\text{Pr}_1, \\ E\left[Y_g^*\right] &= Q_g\text{Pr}_3 + E\left[D - Q_s - \tilde{Q}_r|\Omega_2\right]\text{Pr}_2. \end{aligned}$$

Demand. In Stage 2, the social planner decides for a price p that determines demand D. We obtain the result in Lemma 2 (see Appendix B for a proof) from differentiating welfare with respect

[21] Note that capacity is costly and a planner does not install capacity that lies idle for all possible $\tilde{Q}_r$. In particular, $Q_s > D$ cannot be optimal.

[22] Note that this interval occurs with positive probability if and only if $Q_r > D - Q_s$.

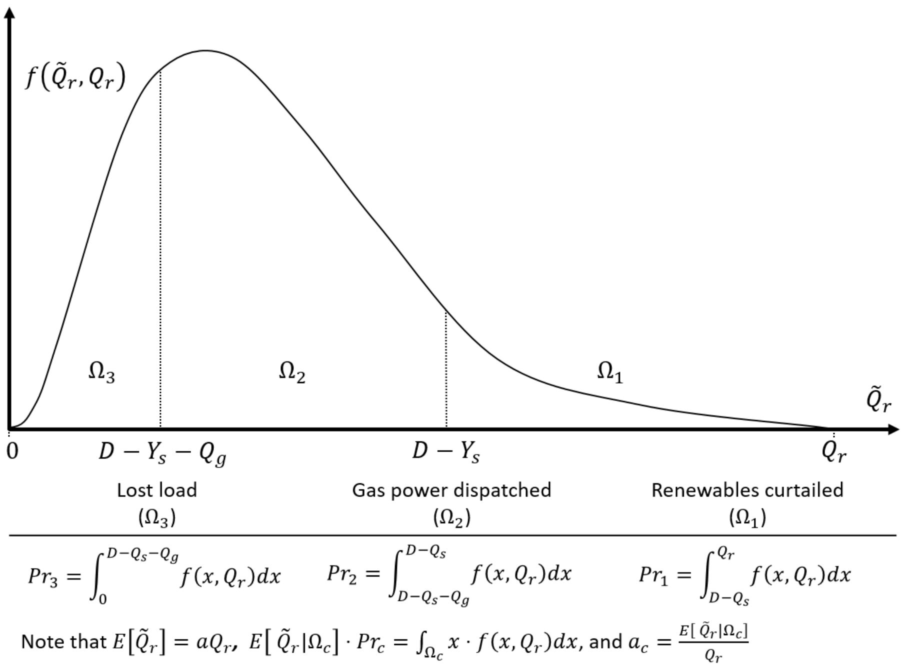

Fig. 2: Illustration of the interval of events

to demand.

Lemma 2. *Suppose that $Q_r^* > 0$. Efficient demand follows from $p = U^{'}(D^*) = c_r \mathrm{Pr}_1 + c_g \mathrm{Pr}_2 + (c_{voll} + \delta)\mathrm{Pr}_3$ with $c_{voll} := c_L + U^{'}(D^*)$.*

Marginal utility at efficient demand, $U^{'}(D^*)$, has both production and consumption components. The production component is given by $c_r \mathrm{Pr}_1$ and $c_g \mathrm{Pr}_2$, that is, the short-run marginal cost of technologies weighted by the respective probabilities that they will be used as marginal technology. Renewables are the marginal technology if Ω_1 realizes, and gas power in the gas power dispatched events (Ω_2). The consumption component considers *curtailment cost* $c_{voll} + \delta$ in lost load events (Ω_3). Curtailment cost comprise that consumers suffer the (marginal) value of lost load (VOLL)—which is the sum of disruption cost and utility losses—and the system operator carries the burden of (marginal) rationing cost for each unit of lost load.

Capacities. Now consider capacity decisions at Stage 1. We maximize welfare (1) w.r.t. Q_s, Q_g, Q_r b. The fact that the random variable is boundedly integrable allows to interchange differentiation and expectation (see Chao, 1983, 2011; Eisenack and Mier, 2019). Taking expectations, yields first-order conditions:

$$\frac{\partial E[W]}{\partial Q_s} = c_r \Pr_1 + c_g \Pr_2 + (c_{voll} + \delta) \Pr_3 - b_s - c_s \le 0 \; [= 0 \text{ if } Q_s^* > 0], \tag{3}$$
$$\frac{\partial E[W]}{\partial Q_g} = (c_{voll} + \delta) \Pr_3 - b_g - c_g \Pr_3 \le 0 \; \left[= 0 \text{ if } Q_g^* > 0\right], \tag{4}$$
$$\frac{\partial E[W]}{\partial Q_r} = (c_g - c_r) \overline{a}_2 \Pr_2 + (c_{voll} + \delta - c_r) \overline{a}_3 \Pr_3 - b_r = 0. \tag{5}$$

Steam power can substitute for renewables in Ω_1 and for gas power in Ω_2 (see first-order condition (3)). Steam power is fully reliable (no stochastic supply) and avoids lost load, i.e., $(c_{voll} + \delta) \Pr_3$ shows benefits from preventing lost load. The sum of these three terms is total benefits and is equal to marginal utility at efficient demand $U^{'}(D^*)$ (see Lemma 2). The last two terms are the LRMC. Thus, a positive level of steam power capacity requires or leads to $b_s + c_s = U^{'}(D^*)$, respectively.

Gas power capacity is not (fully) used in renewables curtailed events Ω_1 (gas power dispatched events Ω_2) and, thus, gas power capacity is beneficial only in lost load events Ω_3. b_g are capacity cost and $c_g \Pr_3$ variable cost during lost load events. Cost occurring in Ω_2 equalize with benefits during that events and we, thus, refrain from showing them in (4). For renewables, benefits and cost are mixed in the first two terms of the first-order condition (5). $c_g, c_{voll} + \delta$ indicate benefits from substituting gas power production in Ω_2 or preventing lost load in Ω_3, respectively, where c_r are related production cost. Similar to gas power in gas power dispatched events (Ω_2), cost and benefits for renewables equalize in renewables curtailed events (Ω_1). Whether or not steam and gas power complements renewables depends on the relative cost. Lemma 3 distinguishes the two possible outcomes (see Appendix C for a proof).

Lemma 3. *Suppose that $Q_r^* > 0$. First-order conditions (3) to (5) define efficient capacities of steam power, gas power, and renewables.*

1. *Suppose that $Q_g^* > 0$ (condition (4) is binding). Efficient capacities of renewables and gas power follow from $\Pr_3^* = \frac{b_g}{c_{voll} + \delta - c_g}$ and $\Pr_2^* = \frac{b_r - (c_{voll} + \delta - c_r)\overline{a}_3 \Pr_3^*}{(c_g - c_r)\overline{a}_2}$.*
2. *Suppose that $Q_g^* = 0$ (condition (4) is not binding). Efficient capacity of renewables follows from $\Pr_3^* = \frac{b_r}{(c_{voll} + \delta - c_r)\overline{a}_3}$.*

We neglected the resulting level of steam power in Lemma 3, but provide some intuition. It is efficient to install a positive level of steam power when $U^{'}(D^*) = b_s + c_s$. When a system with renewables provides $U^{'}(D^*) > b_s + c_s$, we would have $Q_s^* = D_s^*$ and $Q_r^* = 0$. Since we only concentrate on cases with a positive level of renewables, we focus on $U^{'}(D^*) \le b_s + c_s$ in the following. When a system with renewables leads to $U^{'}(D^{*}) = b_s + c_s$, every level $Q_s \in [0, D^{*}]$ is efficient, whereas $U^{'}(D^{*}) < b_s + c_s$ leads to $Q_s^{*} = 0$.

5 Decentralized Solution

We now analyze whether ex-ante or real-time pricing, respectively, allow to decentralize the social planner outcome in Section 4.

5.1 Ex-Ante Pricing

Consider a perfectly competitive market with the same timing described in Figure 1. Firms invest in capacities. The *ex-ante price* p (and demand D) follows from the intersection of inverse demand (i.e., marginal utility at efficient demand) and inverse supply, i.e.,[23]

$$p \;=\; U^{'}(D^{*}) = c_r \Pr_1 + c_g \Pr_2 + (c_{voll} + \delta) \Pr_3. \tag{6}$$

The ex-ante price is a weighted average of the SRMC of the last "technology" used to meet demand. Observe that in lost load events (Ω_3) consumers are the marginal "technology" and the related cost are above the VOLL by rationing cost, δ.

A benevolent planner *enforces* efficient production. Consumers need to pay the resulting price; firms are obligated to deliver the sold amount of electricity. Remember that we use upper-case letters for aggregate values. For firms we use lower-case letters: y_j is production, q_j capacity, and $\tilde{q}_r$ is the realization of the randomly available renewable capacity of a representative firm using the renewable technology. Note that firms have the same expectations of $\tilde{q}_r$, but the final realizations might differ. We assume that there is a continuum of n firms that can decide to enter the market (with technology j) and install q_j or produce y_j, respectively, i.e., $\int_n y_j(n)\,dn = Y_j$, $\int_n q_j(n)\,dn = Q_j$, and $\int_n \tilde{q}_r(n)\,dn = \tilde{Q}_r$. We further assume that (perfectly competitive) firms do not consider how either their own production or their own capacity influence total production, total capacity, or prices. Thus, for firms the occurrence of events Ω_c and the related probabilities $\Pr_c$ are given.

Cost recovery. The decision problem for each firm is to maximize expected profits w.r.t. q_j, measured as the difference between revenues and cost, i.e.,

[23] There is ex-ante demand response, determined by $U^{'} > 0$ with $U^{''} < 0$, but no real-time demand response. Note that ex-ante demand response reflects consumer's reaction to contractual fixed electricity prices or varying day-ahead prices.

$$E\left[\pi_j\right] \;=\; \left(p-c_j\right)E\left[y_j\right]-b_jq_j. \tag{7}$$

Efficiency of a decentralized solution requires that two additional conditions be met. First, profit-maximizing firms must provide efficient capacity levels, which requires that each firm's first-order condition,

$$\frac{\partial E\left[\pi_j\right]}{\partial q_j} \;=\; \left(p-c_j\right)\frac{\partial E\left[y_j\right]}{\partial q_j}-b_j \le 0 \;\left[=0 \text{ if } q_j>0\right], \tag{8}$$

is equivalent to the respective first-order conditions (3) to (5).[24] Second, each firm's expected profit must be zero (*zero-profit condition*); otherwise efficient capacity decisions are not an equilibrium outcome due to exit or entry.

First, suppose that $q_s>0$. The planner enforces $y_s=q_s$. From (8), we obtain $p-b_s-c_s=0$. Substituting in (7) yields $\pi_s=(b_s+c_s-c_s)\,q_s-b_sq_s=0$. Remember from Section 4 that a positive level of steam power capacity in the social planner solution requires that $b_s+c_s=U^{'}\left(D^*\right)$. A price above LRMC of steam power leads to entry of steam power firms until the price drops to b_s+c_s. Thus, a system with steam power only would lead to $p=b_s+c_s$ and is the benchmark price to beat for a system with renewables (and probably gas power). A price strictly below b_s+c_s leads to exit and $q_s=0$, which is the efficient solution given that $p=U^{'}\left(D^*\right)<b_s+c_s$.

Next, suppose that $q_g>0$. The planner enforces $y_g=0$ in Ω_1, $y_g\in(0,q_g)$ so that $Y_g=D-Q_s-\tilde{Q}_r$ in Ω_2, and $y_g=q_g$ in Ω_3. From (7), it follows that $E\left[\pi_g\right]=(p-c_g)\,E\left[y_g\right]-b_gq_g<0$ because $p\le b_s+c_s<b_g+c_g$ and $E\left[y_g\right]<q_g$. An efficient solution that contains gas power capacity cannot be decentralized by ex-ante pricing. The existence of steam power keeps prices below b_s+c_s and, thus, prevents gas power from recovering fixed cost. However, an efficient solution with prices (weakly) below the LRMC of steam power and no gas power capacity might still be decentralized by competitive markets. Proposition 1 shows that this is not the case. For proof see Appendix D.

Proposition 1. *Suppose that $Q_r^*>0$, generating firms sell electricity at the ex-ante price p as specified in (6), and a social planner enforces efficient production as specified in Lemma 1, i.e., $\int_n y_j(n)\,dn=Y_j^*$.*

- *Steam power firms make zero-profits and provide the efficient capacity, i.e., $\int_n q_s(n)\,dn\in[0,D^*]$ for $p=U^{'}\left(D^*\right)=b_s+c_s$ and $q_s=0$ for $p<b_s+c_s$.*

[24] For differentiation, we use the fact that randomly available renewable capacity is assumed to be boundedly integrable, which allows us to interchange differentiation and expectation, as done in Section 4.

- *Gas power firms would suffer losses when entering the market and provide no capacity, i.e.,* $q_g = 0$*.*
- *Given that* $Q_g^* = 0$*, renewables firms provide inefficiently low capacity when entering the market, i.e.,* $\int_n q_r(n)\,dn < Q_r^*$*. For ind, renewables firms make zero-profits for the boundary case of* $E[y_r] = E[\tilde{q}_r]\,\mathrm{Pr}_3$*, profits when* $E[y_r] > E[\tilde{q}_r]\,\mathrm{Pr}_3$*, and losses when* $E[y_r] < E[\tilde{q}_r]\,\mathrm{Pr}_3$*. For corr, renewables firms make profits.*

The ex-ante price is a weighted average of the SRMC of the marginal technology (renewables or gas power) and the marginal cost of curtailment (a combination of consumers and the system operator as marginal technology). Each firm receives the same price no matter what event materializes, that is, there is no *price risk*. In contrast, the market value of electricity generation is low when renewables supply plenty, that is, in the events of excess capacity of renewables. The market value is high when renewables produce little, that is, in the events of lost load.

Steam power firms schedule production before the uncertainty of renewable energies realizes and cannot adjust their schedule according to the final realization, that is, steam power firms do not face any *output risk*. The value of the electricity produced is equal to the ex-ante price. Accordingly, steam power recovers cost (and provides efficient capacities). Gas power firms always adjust their production in response to the stochastic supply of renewable energies and thus face an output risk. In particular, gas power is only fully used in lost load events , where electricity supply is most valuable. The value of produced electricity is above the ex-ante price. In fact, gas power firms require a price above their own LRMC to recover cost, but the actual price is weakly below the LRMC of steam power due to the (potential) presence of steam power in the capacity mix. Thus, gas power cannot exist in a market with ex-ante pricing. Results are ambiguous for renewables, at least for the independence case. For perfectly correlated generating units, the resulting price even leads to profits. Note that positive profits are not an equilibrium outcome, because perfectly competitive firms would enter the market. How to address this issue is subject of the next paragraph.

Support mechanisms. The obvious way to determine optimal support mechanisms is to maximize welfare from a planner's perspective by taking into account the decisions of firms. It is impossible to obtain an analytical tractable solution and interpret the results due to the amount of possible subsidies and taxes. Instead, we simplify the analysis and just aim to decentralize capacity decisions, such that the first-order conditions of profit maximization are equivalent to the first-order conditions of welfare maximization (*efficiency condition*), and firms obtain zero profits to avoid exit and entry (*zero-profit condition*).

The overall goal is to determine support mechanisms that lead to efficient capacity choices.

So, we do not use asterisks in the following because all variables refer to the efficient outcome. For parsimony, we concentrate on the more interesting situation with positive levels of gas power as efficient solution. Note that steam power firms always provide efficient capacities and exactly recover cost. We thus neglect mechanisms for steam power firms in the following.

At this point, our multi-stage decision process changes. Before firms choose capacities in Stage 1, a planner imposes a support mechanism. We consider payments for capacity, σ, and energy supplied (feed-in tariffs or market premia, respectively), τ. Positive payments are a subsidy and negative payments a tax. Subscript j denotes a *technology-specific* payment and subscript $c = 1,2,3$, a *state-contingent feed-in tariff*, e.g., $\tau_{r,2}$ is the feed-in tariff for renewables firms conditional that the interval of events Ω_2 realizes. We have a *flat feed-in tariff* when $\tau_{j,c} = \tau_j$ for all intervals of events.

The expected profits of a representative firm are:

$$E\left[\pi_j\right] = \sum_c \left(p + \tau_{j,c} - c_j\right) E\left[y_j|\Omega_c\right] \mathrm{Pr}_c - \left(b_j - \sigma_j\right) q_j, \tag{9}$$

where $p + \tau_{j,c} - c_j$ are *marginal production profits* and $b_j - \sigma_j$ are *marginal capacity cost* under the chosen support mechanism.

Gas power firms. The production schedule for gas power firms is described above. Taking expectations yields

$$E\left[y_g\right] = q_g \mathrm{Pr}_3 + E\left[y_g|\Omega_2\right] \mathrm{Pr}_2 < q_g. \tag{10}$$

We use this to maximize profits w.r.t. q_g. The efficiency condition of gas power firms is given by

$$\frac{\partial E\left[\pi_g\right]}{\partial q_g} = 0 \quad \Leftrightarrow \quad b_g - \sigma_g = \left(p + \tau_{g,3} - c_g\right) \mathrm{Pr}_3, \tag{11}$$

that is, marginal capacity cost must be equal to expected marginal production profits when lost load (Ω_3) realizes. We use this to obtain the zero-profit condition of gas power firms,

$$E\left[\pi_g\right] = 0 \quad \Leftrightarrow \quad 0 = \left(p + \tau_{g,2} - c_g\right) E\left[y_g|\Omega_2\right] \mathrm{Pr}_2, \tag{12}$$

σ_j	$\tau_{j,1}$	$\tau_{j,2}$	$\tau_{j,3}$	Additional feed-in tariffs
b_j		$c_j - p$		None
None	$c_r - p$	$c_g - p$	$c_{voll} + \delta - p$	$\tau^+_{r,1} = \frac{a_2 q_r}{E[y_r \vert \Omega_1] \Pr_1} \gamma$
$\sigma_g = 0$, $\sigma_r = \frac{a a_2}{a - a_3} \gamma$	$c_r - p$	$c_g - p$	$c_{voll} + \delta - p$	$\tau^+_{r,3} = -\frac{a_2}{(a-a_3) \Pr_3} \gamma$
$\gamma = 0$ for *corr* and $\gamma = \frac{a-a_2}{a_2} \frac{b_r}{a} + \frac{a_2 - a_3}{a_2} (c_{voll} + \delta - c_r) \Pr_3$ for *ind*				

Tab. 1: Examples for optimal support mechanisms

which is fulfilled if and only if $\tau_{g,2} = c_g - p$. Efficiency requires that $\Pr_3 = \Pr_3^*$ (see Lemma 3). One obvious solution is no subsidization of gas power capacity, $\sigma_g = 0$, and a state-contingent feed-in tariff for lost load events equal to the difference of the marginal VOLL and the price ($\tau_{g,3} = c_{voll} - p$, see the second row in Table 1 and note that gas power is not used in Ω_1). Interestingly, the payments necessary to fulfill the efficiency condition $(\sigma_g, \tau_{g,3})$ are independent from those used to fulfill the zero-profit condition $(\tau_{g,2} = c_g - p)$. However, the efficiency condition (11) allows for infinite possibilities. For example, the planner can increase the capacity subsidy and simultaneously reduce $\tau_{g,3}$. This could end up in a situation where capacity is fully subsidized $(\sigma_g = b_g)$ and the feed-in tariff is flat ($\tau_g = c_g - p$, see the first row in Table 1).

Renewables firms. Suppose that a support mechanisms is in place such that gas power firms install efficient capacities. We use efficient production of renewables and additionally account for production in Ω_2 to obtain

$$E[y_r] = E[\tilde{q}_r \vert \Omega_3] \Pr_3 + E[\tilde{q}_r \vert \Omega_2] \Pr_2 + E[y_r \vert \Omega_1] \Pr_1 < E[q_r]. \tag{13}$$

We use this to maximize profits w.r.t. q_r and obtain the efficiency condition of renewables firms,

$$\frac{\partial E[\pi_r]}{\partial q_r} = 0 \quad \Leftrightarrow \quad b_r - \sigma_r = (p + \tau_{r,2} - c_r) \bar{a}_2 \Pr_2 + (p + \tau_{r,3} - c_r) \bar{a}_3 \Pr_3. \tag{14}$$

Similar to the situation for gas power firms, marginal capacity cost must be equal to the expected marginal production profits in the events of full capacity usage (Ω_2, Ω_3). Note that efficiency requires that support mechanisms must be chosen so that $\Pr_2 = \Pr_2^*$ and $\Pr_3 = \Pr_3^*$ (see Lemma 3). We can now substitute for $b_r - \sigma_r$ in (9) by using (14) and obtain the zero-profit condition for renewables firms,

$$E[\pi_r] = 0 \Leftrightarrow 0 \quad = \quad (p + \tau_{r,1} - c_r) E[y_r|\Omega_1] \Pr{}_1 + \tag{15}$$

$$(p + \tau_{r,2} - c_r)(a_2 - \bar{a}_2) q_r \Pr{}_2 + (p + \tau_{r,3} - c_r)(a_3 - \bar{a}_3) q_r \Pr{}_3. \tag{16}$$

Observe that $a_2 q_r = E[y_r|\Omega_2]$ and $a_3 q_r = E[y_r|\Omega_3]$ comes from (9) and $\bar{a}_2 q_r, \bar{a}_3 q_r$ from inserting (14) in (9). Capacity must be fully subsidized ($\sigma_r = b_r$) if the planner chooses a flat feed-in tariff ($\tau_r = c_r - p$), just as was the case for gas power firms (see the first row in Table 1). State-contingent feed-in tariffs fully avoid capacity subsidies, at least for the correlation case. For *corr*, we have $\bar{a}_c = a_c$ so that the second line in (16) vanishes, and payments to fulfill the efficiency condition are independent from those that fulfill the zero-profit condition. The second row in Table 1 shows an obvious solution, which can be proved in a straightforward manner by inserting the supposed payments and tariffs in (14). Appendix E provides additional computations.

For *ind*, in contrast, efficiency and zero-profit conditions are related (see $\bar{a}_2, \bar{a}_3 = a$ in (16)). Consider the same payments as for the correlation case (see second row in Table 1). The payments fulfill the efficiency condition (14), but we obtain $E[\pi_r] = -a_2 q_r \gamma$ when substituting the payments into a renewables firm's profit function. The implementation of an additional feed-in tariff conditional to Ω_1, denoted by $\tau^+_{r,1}$, resolves the problem. $\tau^+_{r,1}$ needs to be chosen so that $\tau^+_{r,1} E[y_r|\Omega_1] \Pr_1 = a_2 q_r \gamma$ (see the last row of the second line in Table 1). Such a payment does not affect the capacity decision of firms, because the capacity constraint does not bind in events of excess capacity (Ω_1). Other state-contingent tariffs distort the efficiency condition and require additional support for capacity. For example, an additional feed-in tariff conditional to Ω_3, denoted by $\tau^+_{r,3}$, needs to be complemented by a capacity payment as shown in the third row in Table 1.[25]

We summarize the results in Proposition 2.

Proposition 2. *Suppose that $Q^*_r, Q^*_g > 0$, generating firms sell electricity at the ex-ante price p as specified by (6), and a social planner enforces efficient production as specified in Lemma 1, i.e., $\int_n y_j(n)\,dn = Y^*_j$. Conditions (11) to (16) define optimal support mechanisms for renewables and gas power firms such that firms make zero profits and provide efficient capacities, i.e., $\int_n q_j(n) = Q^*_j$.*

Obviously, gas power and renewables firms recover cost when total capacity expenses are subsidized and all production profits or losses, respectively, are eliminated by taxing (see the first row in Table 1).[26] Such a support mechanism no longer needs a market and cannot be the target of a market designer. Capacity payments can be reduced when and only when feed-in tariffs

[25] Note that σ_r and $\tau^+_{r,3}$ vary in signs.

[26] Capacity auctions could prevent over-investment when capacity expenses are fully subsidized.

become state-contingent, meaning that tariffs vary according to renewables feed-in. A planner can even fully avoid capacity payments by using payments as indicated in the second row in Table 1. Interestingly, these payments are the same for gas power and renewables firms, at least in the perfect correlation case. Such a setup allows firms to perfectly ensure against their output risk by receiving higher revenues when supply of renewables is scarce (and gas power production high). As soon as generating units of renewables are not perfectly correlated, which is the real-world case, additional payments are needed. A renewables firm now faces uncertain revenues because their own output does not perfectly mirror the output of all renewables firms anymore. For example, a technology-specific additional feed-in tariff ($\tau^+_{r,1}$) or a combination of capacity payments and technology-specific feed-in tariffs ($\sigma^+_r, \tau^+_{r,3}$) restore efficient capacity choices and zero-profits for renewables firms.

5.2 Real-Time Pricing

Up to this point, the analysis has revealed that an ex-ante price will not lead to efficient capacities. Previously, we neglected the possibility that prices adapt after the random variable realizes. Now, we abandon the idea of ex-ante pricing and allow price adjustments in real-time. The aim of the analysis is to investigate whether or not the market could be designed in such a way that support mechanisms are unnecessary. Note that we still not allow for demand adaptation in real-time (dynamic pricing), that is, the assumption that demand is fixed after Stage 2 remains active.

Denote by ρ the *real-time price* and by ρ_c the price that occurs conditional on the interval of events Ω_c that might realize. We define ρ_c as the marginal production cost of the last (and most expensive) technology used to serve demand (also *marginal technology*). Recall that steam power will never be the marginal generating technology because steam power production cannot be reduced (or only at very high cost) or increased after the random variable realizes. So, we need to distinguish between three intervals of events: $\Omega_1, \Omega_2, \Omega_3$. In Ω_1, renewables are the marginal technology because of excess capacity. Therefore, the price equals their SRMC c_r. In Ω_2, gas power is the marginal technology. In Ω_3, a combination of consumers and the system operator acts as marginal technology.[27] We obtain

$$\rho = \begin{cases} c_{voll} + \delta & \text{for } \tilde{Q}_r \in \Omega_3, \\ c_g & \text{for } \tilde{Q}_r \in \Omega_2, \\ c_r & \text{for } \tilde{Q}_r \in \Omega_1, \end{cases} \tag{17}$$

[27] Observe that efficient prices are above the (marginal) VOLL.

where $E[\rho] = c_r \Pr_1 + c_g \Pr_2 + (c_{voll} + \delta) \Pr_3$. Given efficient capacities, production, and demand levels, we obtain $E[\rho] = p = U'(D^*)$. Note that in such a market, the planner no longer needs to enforce production decisions. For example, in Ω_1, the price equals renewable's SRMC and, thus, renewables firms have no incentive to produce more than actually needed. The same applies in Ω_2 for gas power firms. Moreover, in Ω_3, the price equals the marginal cost of curtailment, i.e., $c_{voll} + \delta > c_g > c_r$, which incentivizes production with full capacity by both renewables and gas power firms.

Now turn to investment behavior of firms. Consider the following timing: After capacity decisions of firms in Stage 1, consumers decide on demand in Stage 2. Firms decide production in Stages 3 to 5 and after Stage 5, a price obtains from the intersection of total production and demand. Expected profits of firms are given by

$$E[\pi_j] = \sum_c (\rho_c - c_j) E[y_j | \Omega_c] \Pr_c - b_j q_j. \tag{18}$$

Observe that the real-time price (17) could also be implemented by a support mechanism as described in the second row of Table 1 (without the additional feed-in tariff $\tau_{r,1}^+$). Thus, *investment behavior under real-time pricing is the same as under ex-ante pricing*, with the following support mechanisms: $\tau_1 = c_r - p$, $\tau_2 = c_g - p$, and $\tau_3 = c_{voll} + \delta - p$. We summarize in Proposition 3.

Proposition 3. *Suppose that $Q_r^*, Q_g^* > 0$ and generating firms sell electricity at the real-time price ρ as specified in (17). Firms decide for efficient production. Steam and gas power firms decide in favor of efficient capacities, but renewables firms only for the case of perfect correlation (corr).*

The output risk of gas power firms is resolved by real-time prices that reflect the market value of electricity produced. The proposition further shows that real-time pricing leads to the efficient outcome only under the assumption that marginal generating units of technologies are perfectly correlated. Remember that renewables and gas power firms face an output risk because realized output might be below expected output (e.g., $y_r < E[\tilde{q}_r]$ or $y_g < E[q_g]$). Renewables firms do not face an output risk under perfect correlation: The realized state of the world perfectly mirrors the own output because all renewables firms produce the same. Thus, renewables firms are perfectly ensured against their output risk. When outputs are not perfectly correlated, the realized state of the world is not perfectly correlated to own output (e.g., $\tilde{q}_r = 0$ but $\tilde{Q}_r = \int_n \tilde{q}_r(n)\, dn > D$). Thus, renewables firms cannot (perfectly) ensure against the output risk anymore.

Price decomposition and governmental balance. Why does renewables firms do not recover cost under real-time pricing as soon as marginal generating units are uncorrelated? The difference in the need for additional payments between *ind* and *corr* is primarily founded in the price. We thus decompose the price and explain the origin of γ. First of all, observe that the expected real-time price is equal to the ex-ante price, i.e., $E[\rho] = c_r \Pr_1 + c_g \Pr_2 + (c_{voll} + \delta) \Pr_3 = p$.[28]

Start with the case of perfect correlation. Using $\Pr_2^*$ from Lemma 3, we obtain $E[\rho^{corr}] = \frac{b_r}{a} + c_r + \gamma$. The expected real-time price is equal to the LRMC of renewables plus the *mark-up* γ. Using $E[\rho^{corr}]$ and the first-order conditions (3) to (5), yields expected welfare of $E[W^{corr}] = E[U(X)] - E[\rho^{corr}] E[Y]$. Expected welfare is given by the difference between expected utility from consumption and firms revenues. Note that the zero-profit conditions of firms are fulfilled and, thus, the revenues are equal to the cost (no subsidy scheme is in place). Here, the mark-up is part of the price and thus refers to expected production.

For the independence case, we obtain a price without mark-up, i.e., $E\left[\rho^{ind}\right] = \frac{b_r}{a} + c_r$, but expected welfare contains the mark-up, i.e.,

$$E\left[W^{ind}\right] \quad = \quad E[U(X)] - E\left[\rho^{ind}\right] E[Y] - \gamma E\left[\tilde{Q}_r | \Omega_2\right]. \tag{19}$$

Now, the mark-up refers to expected production of renewables in gas power dispatched events (i.e., $E\left[\tilde{Q}_r | \Omega_2\right]$). The last term in (19) with the mark-up reflects losses of renewables firms, $\int_n E[\pi_r(n)]\, dn = -\gamma E\left[\tilde{Q}_r | \Omega_2\right]$. Subsidy payments from support schemes with $\tau_{r,1}^+$ or $\tau_{r,3}^+, \sigma_r$, respectively, are designed to cover exactly those losses. Consequently, lump-sum taxation of electricity consumers would balance the governmental budget without placing further distortions here. Moreover, this finding holds for all subsidy schemes analyzed in the prior subsection as well.

To summarize, the price is completely sufficient to cover all cost for *corr*, because the *output risk* is reflected in the price by the mark-up γ. For *ind*, the *output risk* is not included in the price and thus renewables firms need additional payments.[29]

Illustrative example with discrete firms. We now have a more intuitive view on the results. Consider a medium sized electricity market with cost for renewables, gas power, and curtailment as follows: $c_r = 0$, $c_g = 50$, and $c_{voll} + \delta = 500$ €/MWh. Assume that demand is 75 GW and 50 GW of gas power is installed. Now consider two renewable firms, each of them owning a wind farm, consisting of 50,000 wind turbines of equal size each (say 1 MW each). The wind farms are either fully available (1), half available (0.5) or not at all available (0). Each of the three events realizes with equal probability, meaning that the expected output of a wind turbine is

[28] The expected price that firms receive is different, e.g., $(c_g E[y_g|\Omega_2] \Pr_2 + (c_{voll} + \delta) E[q_g|\Omega_3] \Pr_3) / E[y_r] > E[\rho]$ is the expected price for gas power firms and sufficient to recover cost of gas power firms.

[29] Note that the only difference in the size of γ is that the endogenous variables $a_2, a_3, \Pr_3$ refer to the respective extreme cases *ind* and *corr*. However, without placing more structure in the probability density function of $\tilde{Q}_r$, we cannot determine whether $E[X] \gtreqless E\left[\tilde{Q}_r | \Omega_2\right]$ nor $\gamma \gtreqless 0$. For example, γ is strictly positive when $\tilde{Q}_r$ is uniformly distributed.

$E\left[\frac{y_r}{q_r}\right]=\frac{1}{3}(1+0.5+0)=0.5$ MWh. Expected output of both wind farms is 50 GWh, whereas maximum output is 100 GWh (and minimum is 0 GWh); meaning that for output greater than 75 GWh we have to curtail renewables (Ω_1), output below 25 GWh leads to loss load (Ω_3), and output between 25 and 75 GWh constitutes gas power dispatched events (Ω_2). For parsimony, we assume that wind turbines within a wind farm are located next to each other and always produce the same (they always face the same weather conditions). However, whether or not both wind farms produce the same depends on the respective case, *ind* or *corr*.

For the correlation case, we obtain three events: $1/1$ (wind farms produce 100 GWh in total), $0.5/0.5$ (50 GWh), and $0/0$ (0 GWh), i.e., $\Omega_1=\{1/1\}$, $\Omega_2=\{0.5/0.5\}$, and $\Omega_3=\{0/0\}$ with $\Pr_1=\Pr_2=\Pr_3=\frac{1}{3}$. The probabilities are optimal when capacity cost are $b_g=90$ and $b_r=8.33$ €/MWh.[30] The expected price is $E\left[\rho^{corr}\right]=\frac{1}{3}(0+50+500)\approx 183.33$ €/MWh.[31] Renewables firms receive revenues only in gas power dispatched events, i.e., $E\left[\frac{\pi_r}{q_r}\right]=\frac{1}{3}\cdot 0.5\cdot 50=8.33$ €/MWh are the revenues per wind turbine (per hour). Note that the revenues per wind turbine exactly cover the capacity cost of renewables.

For the independence case, we obtain nine events and three intervals with $\Pr_1=\Pr_3=\frac{1}{9}$ and $\Pr_2=\frac{7}{9}$.[32] The probabilities are optimal when $b_g=50$ and $b_r=47.22$ €/MWh. The resulting expected price is $E\left[\rho^{ind}\right]=50\cdot\frac{7}{9}+500\cdot\frac{1}{9}\approx 94.44$ €/MWh.[33] Renewables firms receive revenues per turbine (per hour) of $E\left[\frac{\pi_r}{q_r}\right]=\frac{7}{9}\cdot 0.5\cdot 50\approx 19.44$ €/MWh. The revenues are not sufficient to recover capacity cost of 47.22 €/MWh.

Now suppose that a third firm is willing to enter the market to obtain intuition for the different specifications of welfare (see (19)).[34] For the correlation case, the social benefit of adding a turbine is equal to the revenues of already existing firms, because the expected production is 0.5 MWh in gas power dispatched events, but 0 MWh during lost load. For the independence case, an added turbine would produce independently of the two wind farms, that is, with a probability of $\frac{1}{3}$ also during lost load events. The social benefit is $\frac{1}{3}\cdot 1\cdot\left(\frac{7}{9}\cdot 50+\frac{1}{9}\cdot 500\right)+\frac{1}{3}\cdot 0.5\cdot\left(\frac{7}{9}\cdot 50+\frac{1}{9}\cdot 500\right)=$ 47.22 €/MWh—that is, higher than the revenues per turbine—and covers capacity cost exactly. We conclude that, as soon as generating units are (partly) uncorrelated, the social benefit of adding a turbine is higher than the revenues the market supplies because the applied pricing does not ensure against firms' output risk. This effect justifies additional payments for renewables firms even under real-time pricing.

[30] b_g, c_r are the cost of providing a unit of capacity for an hour. We thus abstract from lifetime issues for parsimony.
[31] Observe that the mark-up is $\gamma=E\left[\rho^{corr}\right]-\frac{b_r}{a}-c_r\approx 183.33-\frac{8.33}{0.5}\approx 166.67$ €/MWh.
[32] Intervals are $\Omega_1=\{1/1\}$, $\Omega_2=\{1/0.5, 0.5/1, 1/0, 0/1, 0.5/0.5, 0.5/0, 0/0.5\}$, and $\Omega_3=\{0/0\}$.
[33] Observe that $\frac{b_r}{a}+c_r=\frac{47.22}{0.5}=94.44$ €/MWh.
[34] Remember that firms do not consider their effect on aggregate quantities and prices.

Multi-period model. We used a stylized one-period model to show that real-time pricing alone does not lead to efficient capacity choices of (perfectly competitive) firms due to the correlation of randomly available marginal generating units. Now, we show that this finding is robust in a setting with *periodic demand* and *intermittent renewables*. Changing utility leads to periodic demand and a varying average availability reflects the predictable part of intermittency. Additionally, we keep our assumption of the random renewables supply. Proposition 4 shows that the main finding from the one-period model is robust with respect to multiple periods. Appendix F contains the proof.

Proposition 4. *In a setting with real-time pricing, periodic demand, predictable intermittency, and stochastic intermittency, renewables firms decide for efficient capacities when and only when the availabilities of their marginal generating units are perfectly correlated (corr).*

The proposition underlines that the correlation of marginal generating units of renewables is crucial for future market design not only in a stylized one-period model, but also with respect to closer real-world settings.

5.3 Consumer Behavior

We now add the last piece to analyze existing market designs in the next section. In the analysis thus, we have ignored final consumers' incentives to actually decide for efficient demand. Suppose that a consumer decides for demand Δ. $y \leq \Delta$ is the served demand, so that $x = \min\{\Delta, y\}$ is final consumption. A consumer obtains utility u from consumption with $u^{'} > 0$, $u^{''} < 0$. We assume that there is a continuum of m heterogeneous consumers so that $\int_m \Delta(m)\, dm = D$, $\int_m y(m)\, dm = Y$, and $\int_m u(m, \Delta(m))\, dm = U(D)$. We further assume that consumers do not consider how their own decisions for Δ influences total demand, prices, or the occurrence of events Ω_c and the related probabilities Pr_c. It is straightforward to show that maximizing welfare requires that each consumer schedules demand until $u^{'}(\Delta) = c_r \mathrm{Pr}_1 + c_g \mathrm{Pr}_2 + (c_{voll} + \delta)\mathrm{Pr}_3$. Thus, marginal utility and also marginal cost of curtailment are the same for each consumer. In particular, we have $u^{'}(\Delta) = U^{'}(D)$ and $c_{voll} + \delta = c_L + u^{'}(\Delta) + \delta = c_L + U^{'}(D) + \delta$ (see Lemma 2).

Suppose that a support mechanism is in place that leads to efficient capacity choices of firms (in Stage 1). In case of lost load, the system operator curtails consumers with the lowest willingness-to-pay and pays compensation τ_L for each unit of lost load. Note that not all consumers get curtailed in case of lost load, and the curtailed ones not necessarily consume zero electricity. $\mathrm{Pr}_L \in [0,1]$ denotes the individual probability of each consumer of getting curtailed in case of lost load.

Consumers pay ex-ante prices. Presume that the ex-ante price follows from the intersection of inverse demand and supply, i.e., $p = u^{'}(\Delta) = c_r \mathrm{Pr}_1 + c_g \mathrm{Pr}_2 + (c_{voll} + \delta)\mathrm{Pr}_3$, in Stage 2. In

Stages 3 to 5, the planner enforces efficient production. Demand Δ follows from maximizing (expected) consumer surplus, which is given by

$$
\begin{aligned}
E[cs] &= (u(\Delta) - p\Delta)\Pr{}_{12} \\
&\quad + [(u(\Delta) - p\Delta)(1 - \Pr{}_L) + (u(y) - py - (c_L - \tau_L)(\Delta - y))\Pr{}_L]\Pr{}_3. \qquad (20)
\end{aligned}
$$

The first line in (23) represents consumer surplus when there is no lost load. The second line shows consumer surplus in case of lost load. Non-curtailed consumers consume Δ, whereas curtailed ones consume $y < \Delta$. From differentiating $E[cs]$ w.r.t. Δ, we obtain:

$$
\begin{aligned}
\frac{\partial E[cs]}{\partial \Delta} &= \left(u^{'}(\Delta) - p\right)\Pr{}_{12} \\
&\quad + \left[\left(u^{'}(\Delta) - p\right)(1 - \Pr{}_L) - (c_L - \tau_L)\Pr{}_L\right]\Pr{}_3 \leq 0\ [= 0 \text{ if } \Delta > 0]. \qquad (21)
\end{aligned}
$$

From the binding first-order condition (21), it follows that

$$
u^{'}(\Delta) = p + (c_L - \tau_L)\frac{\Pr{}_L \Pr{}_3}{\Pr{}_{12} + (1 - \Pr{}_L)\Pr{}_3}. \qquad (22)
$$

Observe that consumers with $\Pr_L > 0$ would not decide for efficient demand. A compensation of $\tau_L = c_L$ corrects this market failure. The planner needs to account for disruption cost that arise from lost load, because these cost are not fully internalized in the price for curtailed consumers. We summarize in Proposition 5.

Proposition 5. *Suppose that $Q_r^*, Q_g^* > 0$, final consumers buy electricity at the ex-ante price p as specified by (6), a social planner enforces efficient production as specified in Lemma 1, and implements support mechanisms for renewables and gas power firms as defined by conditions (11) to (16). Final consumers decide for the efficient demand, i.e., $\int_m \Delta(m)\,dm = D^*$, when curtailed consumers receive compensation of c_L for each unit of lost load.*

Final consumers do consider utility losses from lost load but not the disruption cost. Intuitively, the ability to curtail consumers when production capability is not sufficient to cover the whole demand is similar to an additional production technology. Offering system flexibility requires—in the same way as does gas power—compensation or, more precise, a subsidy. However, the subsidy paid to consumers does not contain marginal rationing cost. Consumers do not prevent curtailment and the occurrence of related rationing cost, whereas firms do.[35]

[35] An reactive consumer who reduces its demand in response to lost load events prevents rationing cost and must receive the same payments as the firms.

Consumers pay real-time prices. Recall from Propositions 2 and 3 that generation and investment behavior of firms is the same under ex-ante pricing with a specific support mechanism ($\tau_1 = c_r - p$, $\tau_2 = c_g - p$, and $\tau_3 = c_{voll} + \delta - p$) and real-time pricing. However, consumer behavior changes fundamentally under real-time pricing. Suppose that τ_L^{rtp} is the compensation for curtailed consumers in case of lost load. The expected consumer surplus changes to

$$\begin{aligned} E[cs] &= (u(\Delta) - c_r\Delta)\Pr_1 + (u(\Delta) - c_g\Delta)\Pr_2 \\ &+ \left[(u(\Delta) - (c_{voll}+\delta)\Delta)(1-\Pr_l) + \left(u(y) - (c_{voll}+\delta)y - \left(c_L - \tau_L^{rtp}\right)(\Delta - y)\right)\Pr_L\right]\Pr_3 \end{aligned} \tag{23}$$

Note that real-time prices are now changing with the realization of Ω_1, Ω_2, or Ω_3, respectively. From differentiating $E[cs]$ w.r.t. Δ and solving the binding first-order condition for $u'(\Delta)$, we obtain:

$$u'(\Delta) = E[\rho] - \left(\delta + \tau_L^{rtp}\right)\Pr_L \Pr_3, \tag{24}$$

where $E[\rho] = c_r\Pr_1 + c_g\Pr_2 + (c_{voll}+\delta)\Pr_3$. Surprisingly, the compensation for lost load (to constitute efficient demand) is indeed a tax, i.e., $\tau_L^{rtp} = -\delta$. Consumers do not need compensation for disruption cost, because disruption cost are already fully internalized in the price (which is $c_{voll} + \delta$ in Ω_3). In turn, curtailed consumers need to pay for rationing cost. We summarize in Proposition 6.

Proposition 6. *Suppose that $Q_r^*, Q_g^* > 0$, final consumers buy electricity at the real-time price ρ as specified by (17), and a social planner imposes $\sigma_r, \tau_{r,3}^+$ as specified in the last line of Table 1. Final consumers decide for efficient demand, i.e., $\int_m \Delta(m)\,dm = D^*$, when curtailed consumers get taxed at δ for each unit of lost load.*

Proposition 3 shows a counter-intuitive pricing structure for final consumers under real-time pricing. The price in lost load events is equal to the cost of curtailment, that is, the sum of disruption cost, rationing cost, and utility losses from curtailment. However, the marginal consumer (the last consumer served) is willing to pay for disruption cost and utility losses only, i.e., $c_L + u'(\Delta) < c_{voll} + \delta$. The social planner needs to impose opportunity cost of not consuming electricity by taxing curtailed consumers at rationing cost δ to incentivize efficient demand.

6 Market Design

We do not observe designs consisting of single ex-ante or real-time markets in real-world power markets. Indeed, we can differentiate between two major philosophies (see Wilson, 2002; Cramton, 2017; Wolak et al., 2020). First, there is a combination of a day-ahead with a real-time market that finally settles delivery pledges. Trading among markets is generally possible. Such a combination is often applied in Northern America. European markets in turn combine day-ahead and intraday markets with balancing markets that finally settle differences between ex-ante delivery pledges and real-time supply. The difference of the European balancing to the Northern American real-time market is usually the pricing system (capacity charges or energy-only, respectively) and the resulting trading volume, which are generally lower in balancing markets, leading to kind of poorer pricing (Cramton, 2017).

Moreover, both philosophies use different forms of undersupply penalties when real-time generation is below the ex-ante delivery pledge (e.g., Sunar and Birge, 2019). We implement such penalties as soon as a firms cannot fulfill their ex-ante pledge and lost load occurs in real time. Some markets (e.g., United Kingdom and Sweden) allow for consumer compensation when the (distribution) system operators fails in restoring electricity supply within a certain time frame, whereas the origin of such a black out neglects, for example, extreme weather events. There are no markets that use consumer compensation schemes in case lost load results from *bad* ex-ante delivery pledges or insufficient provision of reliability by means of secured generation capacity, respectively. However, Subsection 5.3 shows that efficient demand—when final consumers are not reactive in real time—requires to compensate consumers when they pay ex-ante prices and tax consumers when they pay real-time prices. In light of searching for an efficient outcome, we modify all analyzed market structures by such a consumer compensation/tax scheme. The organizational overhead is low because undersupply from firms is already measured and accounted for by means of the undersupply penalty. Such undersupply penalty is used to carry cost of the system operator (rationing cost) and consumer compensation (disruption cost) when consumer pay ex-ante prices. The social planner just need to transfer payments ex-post from generators to curtailed consumers. Another possible compensation scheme neglects undersupply penalties and directly places the burden of compensation to firms. However, such a scheme seems to come with tremendous overhead because it might be efficient to curtail consumers of some firms that do not deviate from their ex-ante delivery pledge, because those have the lowest WTP. As a consequence, firms that fulfill their pledges but curtail consumers need to get compensated again from firms that do not fulfill pledges. We decide to use the simpler first metric in the following and place the system operator between firms and consumers, that is, firms pay undersupply penalties to the system operator (as it is the case already) and consumers receive compensation from the system operator (as it is not implemented yet).

Consumer compensation is not necessary for obtaining efficient demand when consumers pay real-time prices. Indeed, consumers need to get taxed at rationing cost. The high real-time price in combination with the tax in lost load events already prevents inefficient demand decisions. Those taxes can be obtained from final consumers via regular network changes and, thus, it is again the system operator that serves as clearing entity.

6.1 Northern American Design

The Northern American markets follow the integrated approach of centrally optimized markets combining a day-ahead market (DAM) for unit commitment and scheduling with a real-time market (RTM) for security-constrained economic dispatch (Cramton, 2017). We now show how such a combination impacts efficient pricing.

Consumers pay day-ahead price. Suppose that generating firms and final consumers interact on a day-ahead market.[36] The outcome is demand D and the day-ahead price p. A representative firm needs to deliver Δ_j (so that $\sum_j \Delta_j = D$). Firms also engage on a real-time market where they can purchase ($y_{j,rtm} < 0$) or offer ($y_{j,rtm} > 0$) electricity. The outcome is the real-time price as specified by (17). Firms need to pay the undersupply penalty τ_P and rationing cost δ to system operators if they are not able to deliver Δ_j either by own generation or purchases.[37] Profits of a representative firm using technology j are

$$\begin{aligned} E\left[\pi_j\right] &= p\left(\Delta_j \mathrm{Pr}_{12} + E\left[y_j - y_{j,rtm} | \Omega_3\right] \mathrm{Pr}_3\right) + \sum_c \rho_c E\left[y_{j,rtm} | \Omega_c\right] \mathrm{Pr}_c \\ &\quad - (\tau_P + \delta) E\left[\Delta_j - y_j + y_{j,rtm} | \Omega_3\right] \mathrm{Pr}_3 - b_j q_j - c_j E\left[y_j\right]. \end{aligned} \tag{25}$$

Firms earn money from selling on the day-ahead market (first term in the first line). In Ω_1, Ω_2, firms are able to deliver Δ_j, i.e., $y_{j,rtm} = y_j - \Delta_j$. In Ω_3, firms deliver less then Δ_j when $y_j - y_{j,rtm} < \Delta_j$. Firms receive profits when selling but have additional cost when purchasing on the real-time market (second term in the first line). Further cost occur from the undersupply penalty, rationing cost, capacity investments, and electricity generation (second line). We can differentiate (25) w.r.t. Δ_j to obtain the first-order condition of a firm's profit maximization problem on the day-ahead market, i.e.,

[36] Remember that we neglect the role of intermediary retailers that close the gap between firms—that engage on wholesale markets—and final consumers—that usually have contracts with intermediaries—for most parts of this article because the outcome is equivalent as shown in Subsection 6.3.

[37] System operators use τ_P to compensate curtailed consumers.

$$\frac{\partial E\left[\pi_j\right]}{\partial \Delta_j} = p\,\mathrm{Pr}_{12} - c_r\,\mathrm{Pr}_1 - c_g\,\mathrm{Pr}_2 - (\tau_P + \delta)\,\mathrm{Pr}_3 \leq 0 \,\left[= 0 \text{ if } \Delta_j > 0\right]. \tag{26}$$

We can rearrange binding (26) and obtain $p = U(D^*)$ when $\tau_P = c_L = \tau_L^*$. Proposition 7 summarizes.

Proposition 7. *Suppose that $Q_r^*, Q_g^* > 0$, demand and price follows from the interaction of generating firms and final consumers on a day-ahead market, and generating firms trade electricity at the real-time price ρ as specified by (17) on a real-time market. The resulting day-ahead demand and day-ahead price are efficient when generating firms pay c_L for undersupply, generating firms carry the burden of rationing cost δ, and curtailed consumers receive c_L for each unit of lost load.*

Proposition 7 underlines that efficient pricing requires undersupply penalties in lost load events (at disruption cost). Moreover, the integrated DAM-RTM presented here is clear about payments streams. Generating firms are obliged to deliver a certain amount of demand (by day-ahead delivery pledges). Firms pay undersupply penalties to the system operator when not doing so. The system operator in turn covers own (rationing) cost and compensates consumers that suffer from lost load. More precisely, such consumer compensation addresses the reliability externality described in Wolak et al. (2020).

Consumers pay real-time price. Some Northern American day-ahead markets just fix the demand but final prices establish in real-time markets. Profits of a representative firm change to

$$\begin{aligned} E\left[\pi_j\right] &= \Delta_j\left(\rho_1\,\mathrm{Pr}_1 + \rho_2\,\mathrm{Pr}_2\right) + \rho_3 E\left[y_j - y_{j,rtm}|\Omega_3\right]\mathrm{Pr}_3 + \sum_c \rho_c E\left[y_{j,rtm}|\Omega_c\right]\mathrm{Pr}_c \\ &\quad - (\tau_P + \delta)\,E\left[\Delta_j - y_j + y_{j,rtm}|\Omega_3\right]\mathrm{Pr}_3 - b_j q_j - c_j E\left[y_j\right]. \end{aligned} \tag{27}$$

Firms still earn money from selling on the day-ahead market but now real-time prices apply (see first and second term in the first line). Differentiation w.r.t. Δ_j yields

$$\frac{\partial E\left[\pi_j\right]}{\partial \Delta_j} = -(\tau_P + \delta)\,\mathrm{Pr}_3 \leq 0 \,\left[= 0 \text{ if } \Delta_j > 0\right]. \tag{28}$$

Obviously, day-ahead demand is efficient when firms receive δ, the opposite of carrying the burden of rationing cost. Proposition 8 summarizes.

Proposition 8. *Suppose that $Q_r^*, Q_g^* > 0$, demand follows from the interaction of generating firms and final consumers on the day-ahead market, generating firms trade electricity at the real-time* *price ρ as specified by (17) on a real-time market, and final consumers pay the real-time price. The resulting day-ahead demand is efficient when generating firms do not pay for undersupply and rationing cost, and curtailed consumers get taxed at δ for each unit of lost load.*

No undersupply penalties (in lost load events) are necessary when final consumers pay real-time prices (see Proposition 6). Indeed, curtailed consumers pay rationing cost (see Proposition 6) to the system operator to cover rationing expenses. Note that most models (and real-world power markets) abstract from rationing cost or those cost are quite low, respectively. However, designs that deliver such unexpected outcome might be unwanted again. Additionally, most final consumers are not willing to pay real-time prices.

Price cap and capacity payments. Some Northern American markets additional implement capacity payments (or markets) to ensure long-run resource adequacy. A price cap that cuts real-time prices below $c_{voll} + \delta$ is the only way of justifying long-run capacity payments in markets providing a real-time signal for generating firms in our model setup.[38] Consider the setting where consumers pay day-ahead (and not real-time) prices. Suppose that $\overline{\rho}_c$ is the capped real-time price with $\overline{\rho}_1 = c_r$, $\overline{\rho}_2 = c_g$, and $\overline{\rho}_3 < c_{voll} + \delta$. Additionally, firms receive σ_j for each capacity unit provided and suffer the undersupply penalty $\tau_{j,3}$ when they are not able to deliver Δ_j.

Profits of a firm change to

$$\begin{aligned} E\left[\pi_j\right] &= p\left(\Delta_j \Pr{}_{12} + E\left[y_j - y_{j,rtm}|\Omega_3\right]\Pr{}_3\right) + \sum_c \overline{\rho}_c E\left[y_{j,rtm}|\Omega_c\right]\Pr{}_c \\ &\quad - \left(\tau_{j,3} + \delta\right) E\left[\Delta_j - y_j + y_{j,rtm}|\Omega_3\right]\Pr{}_3 - \left(b_j - \sigma_j\right) q_j - c_j E\left[y_j\right]. \end{aligned} \tag{29}$$

Now firms earn (pay) less for trading electricity on the real-time market in lost load events (see $\overline{\rho}_c$ in the second term of the first line) but additionally receive the capacity payment. Differentiation w.r.t. Δ_j yields the same result regarding efficient pricing as in (26). However, now firms (zero) profits and capacity decisions does not follow directly from Subsections 5.1 and 5.2.

Efficient steam power dispatch is $y_s = q_s$ in all events. Steam power firms sell $\Delta_s \geq y_s = q_s$ ex-ante because they cannot react in real-time, i.e., $y_{s,rtm} \leq 0$. Profits of steam power firms are

[38] Others are risk-avers firms, poor pricing, and market power that prevents new competitors from entering the market.

$$
\begin{aligned}
E\left[\pi_s\right] &= p\Delta_s \Pr{}_{12} + pE\left[q_s - y_{s,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad + c_r E\left[q_s - \Delta_s|\Omega_1\right]\Pr{}_1 + c_g E\left[q_s - \Delta_s|\Omega_2\right]\Pr{}_2 + \overline{\rho}_3 E\left[y_{s,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad - \left(\tau_{s,3} + \delta\right) E\left[\Delta_s - q_s + y_{s,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad - \left(b_s - \sigma_s\right) q_s - c_s q_s. \qquad (30)
\end{aligned}
$$

The first line represents revenues from selling day-ahead, the second line are costs from buying in real-time, the third line are additional cost from potential undersupply and rationing, and the last line are capacity and production cost less capacity payments.

Remember that the efficient gas power firm dispatch is $y_g = 0$ in Ω_1, $y_g < q_g$ in Ω_2, and $y_g = q_g$ in Ω_3. Moreover, we have $y_{g,rtm} = -\Delta_g$ in Ω_1, $y_{g,rtm} = y_g - \Delta_g$ in Ω_2, and $y_{g,rtm} < q_g - \Delta_g$ in Ω_3. Gas power firms profits are

$$
\begin{aligned}
E\left[\pi_g\right] &= p\Delta_g \Pr{}_{12} + pE\left[q_g - y_{g,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad - c_r \Delta_g \Pr{}_1 + c_g E\left[y_g - \Delta_g|\Omega_2\right]\Pr{}_2 + \overline{\rho}_3 E\left[y_{g,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad - \left(\tau_{g,3} + \delta\right) E\left[\Delta_g - q_g + y_{g,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad - \left(b_g - \sigma_g\right) q_g - c_g\left(E\left[y_g|\Omega_2\right]\Pr{}_2 + q_g \Pr{}_3\right). \qquad (31)
\end{aligned}
$$

Renewables firm dispatch is $y_r < \tilde{q}_r$ in Ω_1, $y_r = \tilde{q}_r$ and $y_{r,rtm} = \tilde{q}_r - y_r$ in Ω_2, Ω_3. Renewables firms profits are

$$
\begin{aligned}
E\left[\pi_r\right] &= p\Delta_r \Pr{}_{12} + pE\left[\tilde{q}_r - y_{r,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad + c_r E\left[y_r - \Delta_r|\Omega_1\right]\Pr{}_1 + c_g E\left[\tilde{q}_r - \Delta_r|\Omega_2\right]\Pr{}_2 + \overline{\rho}_3 E\left[y_{r,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad - \left(\tau_{r,3} + \delta\right) E\left[\Delta_r - \tilde{q}_r + y_{r,rtm}|\Omega_3\right]\Pr{}_3 \\
&\quad - \left(b_r - \sigma_r\right) q_r - c_r\left(E\left[y_r|\Omega_1\right]\Pr{}_1 + E\left[\tilde{q}_r|\Omega_{23}\right]\Pr{}_{23}\right). \qquad (32)
\end{aligned}
$$

We now derive profit maximizing capacity decisions for firms. We first determine resulting capacity levels by differentiating (31) w.r.t. q_j, i.e.,

$$\frac{\partial E[\pi_s]}{\partial q_s} = c_r \Pr{}_1 + c_g \Pr{}_2 + (\tau_{s,3} + p + \delta) \Pr{}_3 - b_s + \sigma_s - c_s \leq 0 \ [= 0 \text{ if } q_s > 0], \quad (33)$$

$$\frac{\partial E[\pi_g]}{\partial q_g} = \left(\tau_{g,3} + p + \delta\right) \Pr{}_3 - b_g + \sigma_g - c_g \Pr{}_3 \leq 0 \ [= 0 \text{ if } q_g > 0], \quad (34)$$

$$\frac{\partial E[\pi_r]}{\partial q_r} = (c_g - c_r) \overline{a}_2 \Pr{}_2 + (\tau_{r,3} + p + \delta - c_r) \overline{a}_3 \Pr{}_3 - b_r + \sigma_r \leq 0 \ [= 0 \text{ if } q_r > 0]. \quad (35)$$

Using Lemma 3 for $Q_g^* > 0$ and binding first-order conditions (33) to (35) allows to derive $\sigma_j, \tau_{j,3}$ combinations that lead to efficient procurement of capacity, i.e., $\Pr_2 = \Pr_2^*$ and $\Pr_3 = \Pr_3^*$. We obtain $\sigma_s = (c_{voll} - p - \tau_{s,3}) \Pr_3$, $\sigma_g = \left(c_{voll} - p - \tau_{g,3}\right)$, and $\sigma_r = (c_{voll} - p - \tau_{r,3}) \overline{a}_3 \Pr_3$. One obvious solution is to set the undersupply penalty equal to $c_{voll} - p$ so that there is no capacity payment. The outcome would be as described in Proposition 7. For parsimony, we concentrate on the simpler case of perfect correlation for the following analysis, i.e., $\overline{a}_3 = a_3$. We obtain

$$E\left[\pi_j\right] = \Delta_j \left(c_{voll} - p - \tau_{j,3}\right) \Pr{}_3 + E\left[y_{j,rtm} | \Omega_3\right] \left(\overline{\rho}_3 - p - \tau_{j,3} - \delta\right) \Pr{}_3 \quad (36)$$

from inserting σ_j in (30), (31), or (32), respectively. Solving $E\left[\pi_j\right] = 0$ for $\tau_{j,3}$ and using this to determine σ_j yields

$$\tau_{j,3} = \frac{(\rho_3 - \delta) \Delta_j + (\overline{\rho}_3 - \delta) E\left[y_{j,rtm} | \Omega_3\right]}{\Delta_j + E\left[y_{j,rtm} | \Omega_3\right]} - p, \quad (37)$$

$$\sigma_s = (\rho_3 - \overline{\rho}_3) \frac{E\left[y_{s,rtm} | \Omega_3\right]}{\Delta_s + E\left[y_{s,rtm} | \Omega_3\right]} \Pr{}_3 \quad (38)$$

$$\sigma_g = (\rho_3 - \overline{\rho}_3) \frac{E\left[y_{g,rtm} | \Omega_3\right]}{\Delta_g + E\left[y_{g,rtm} | \Omega_3\right]} \Pr{}_3, \quad (39)$$

$$\sigma_r = (\rho_3 - \overline{\rho}_3) \frac{E\left[y_{r,rtm} | \Omega_3\right]}{\Delta_r + E\left[y_{r,rtm} | \Omega_3\right]} a_3 \Pr{}_3. \quad (40)$$

Proposition 9 summarizes.

Proposition 9. *Suppose that $Q_r^*, Q_g^* > 0$, demand and price follows from the interaction of generating firms and final consumers on a day-ahead market, generating firms trade electricity at the capped real-time price $\overline{\rho}_1 = c_r$, $\overline{\rho}_2 = c_g$, and $\overline{\rho}_3 < c_{voll} + \delta$ on a real-time market, and generating firms receive a capacity payment σ_j and carry the burden of an undersupply penalty $\tau_{j,3}$ when they do not fulfill their day-ahead delivery pledge. Conditions (37) to (40) define optimal capacity payments and undersupply penalties for generating firms such that firms make zero profits and* *provide efficient capacities, i.e.,* $\int_n q_j(n) = Q_j^*$.

Observe that day-ahead demand D would be distorted when $\tau_{j,3} \neq c_L$. The only scheme that leads to the efficient outcome follows from differentiating between generating firms that engage on the day-ahead market (with final consumers) and those that decide to work on the real-time market only. Generating firms with day-ahead obligations would not *plan* to sell (buy) structurally on the real-time market, i.e., $E\left[y_{j,rtm}\right] = 0$, although they are able to balance own excess or shortfall. Those firms need to pay an undersupply penalty in case of lost load, i.e., $\tau_{j,3} = \tau_L^* = c_L$, and are not further impacted by the price cap. Generating firms without day-ahead obligations, i.e., $\Delta_j = 0$, can trade their entire production on the real-time market. They receive $\sigma_s = \sigma_g = (\rho_3 - \overline{\rho}_3)\Pr_3$ or $\sigma_r = (\rho_3 - \overline{\rho}_3)\, a_3 \Pr_3$, respectively. It would make sense that all renewables and steam power firms engage on the day-ahead market and cover entire demand of final consumers. Those firms would carry the burden of undersupply penalties and rationing cost (in case of lost load). Gas power firms in turn do not (or partially do not) engage here but on the real-time market instead. Their higher flexibility makes them predestined for balancing shortfalls in renewables production. Then, the entire capacity payments (or market) is only for gas power firms. Indeed, long-run resource adequacy is ensured by high enough back-up (gas power) capacity to balance production shortfalls (from other generators).

6.2 European Design

The European design follows the trading approach with combinations of day-ahead and intraday markets, where the final deviations are settled by means of a balancing market (Cramton, 2017). The assumptions of our model make differentiation between day-ahead and intraday markets obsolete because the uncertainty realizes in real time. However, we conduct a more detailed analysis of multiple ex-ante markets in Subsection 6.4. Now, we concentrate on the role of the balancing market in adjusting the outcome of a single ex-ante (e.g., day-ahead) market. Suppose that firms supply final consumers directly and consumers pay ex-ante prices that are contractual fixed, as it is the case in forward, day-ahead, and intraday markets or when electricity is traded over-the-counter. A representative firm signs contracts with consumers of Δ_j at the ex-ante price p in advance of actual delivery. Firms purchase *imbalance energy*—offered by other firms—or pay an undersupply penalty τ_P if they are not able (or not willing) to produce Δ_j on their own. Conversely, firms offer imbalance energy in the case they experience excess capacity. $y_{j,im} > 0$ is imbalance energy sold to other firms, where $y_{j,im} < 0$ is purchased imbalance energy.

Energy-only balancing market. Suppose that firms do not receive payments for providing capacity. Renewables and gas power firms offer imbalance energy in total when there is excess capacity of renewables (Ω_1). Renewables are the marginal technology, and the price for imbalance energy is c_r. Note that firms do not need to be symmetric with regard to the availability of renewable capacity. For example, some renewables firms might experience excess capacity and offer imbalance energy, whereas other renewable firms demand imbalance energy. What matters here is solely that $\int_n \tilde{q}_r(n)\,dn = \tilde{Q}_r \in \Omega_1$. Also in gas power dispatched events (Ω_2), renewables and gas power firms offer imbalance energy, but the offering from renewables firms is fully used so that gas power is the marginal technology. The resulting imbalance energy price is c_g. In case of lost load (Ω_3), there is not enough imbalance energy to serve demand, firms pay the undersupply penalty, and the system operator needs to decide to curtail some consumers. Note that curtailment of consumers demands for compensation and firms need to carry the burden of rationing cost. Additionally accounting for utility losses of lost load, the marginal cost of curtailment $c_{voll} + \delta$ is the price for imbalance energy in Ω_3. The result is the real-time price ρ as described by (17). Then, profits of a representative firm using technology j are

$$\begin{aligned} E\left[\pi_j\right] &= p\left(\Delta_j \Pr{}_{12} + E\left[y_j - y_{j,im}|\Omega_3\right]\Pr{}_3\right) + \sum_c \rho_c E\left[y_{j,im}|\Omega_c\right]\Pr{}_c \\ &\quad - (\tau_P + \delta)\, E\left[\Delta_j - y_j + y_{j,im}|\Omega_3\right]\Pr{}_3 - b_j q_j - c_j E\left[y_j\right]. \end{aligned} \tag{41}$$

Firms earn money from selling to consumers (first term in the first line). In Ω_1, Ω_2, firms are definitely able to sell the contracted amount Δ_j. In Ω_3, they pay the undersupply penalty when $y_j - y_{j,im} < \Delta_j$. Moreover, firms sell (buy) imbalance energy to (from) other firms (second term in the first line). Additional cost occur from the undersupply penalty, rationing cost, capacity investments, and production.

In Ω_1 and Ω_2, total production is sufficient to serve demand so that $y_{j,im} = y_j - \Delta_j$. Using this, we can differentiate (41) w.r.t. Δ_j to obtain the first-order condition of a firm's profit maximization problem:

$$\frac{\partial E\left[\pi_j\right]}{\partial \Delta_j} = p\Pr{}_{12} - c_r \Pr{}_1 - c_g \Pr{}_2 - (\tau_P + \delta)\Pr{}_3 \leq 0 \ \left[= 0 \text{ if } d_j > 0\right]. \tag{42}$$

Proposition 10 summarizes the results.

Proposition 10. *Suppose that $Q_r^*, Q_g^* > 0$, generating firms sign contracts with consumers to deliver Δ_j at the ex-ante price p, and trade electricity at the real-time price ρ as specified by (17) in an* energy-only *balancing market. The resulting ex-ante price and demand is efficient when generating firms pay c_L for undersupply, carry the burden of rationing cost δ, and final consumers receive c_L for each unit of lost load.*

Note that the European design with energy-only balancing markets providing real-time price signals and the Northern American design with day-ahead and real-time markets is equivalent in our model. However, European markets tend to have intraday markets that reduce the balancing market volume. This structure leads to prices that depend on a thinner market and thus might be not as efficient as powerful real-time markets in Northern America (Cramton, 2017).

Balancing market with capacity payments. Most balancing markets are not working as energy-only markets but mainly pay firms for providing capacity. We thus introduce a capacity payment (σ_j), an additional (technology-specific) feed-in tariff for both energy traded ex-ante (τ_j) and those traded as imbalance energy in real-time ($\tau_{j,im}$). The profits of a firm become

$$\begin{aligned} E\left[\pi_j\right] &= \left(p+\tau_j\right)\left(\Delta_j \Pr{}_{12} + E\left[y_j - y_{j,im}|\Omega_3\right]\Pr{}_3\right) + \tau_{j,im}\sum_c E\left[y_{j,im}|\Omega_c\right]\Pr{}_c \\ &\quad -\left(\tau_P+\delta\right)E\left[\Delta_j - y_j + y_{j,im}|\Omega_3\right]\Pr{}_3 - \left(b_j-\sigma_j\right)q_j - c_j E\left[y_j\right]. \end{aligned} \tag{43}$$

Differentiation yields

$$\frac{\partial E\left[\pi_j\right]}{\partial \Delta_j} = -\left(\tau_P+\delta\right)E\left[\Delta_j - y_j + y_{j,im}|\Omega_3\right]\Pr{}_3. \tag{44}$$

It is straightforward to show that such situation is equivalent to one described in the first line of Table 1. We thus summarize in Proposition 11.

Proposition 11. *Suppose that $Q_r^*, Q_g^* > 0$, demand $\sum_j \Delta_j = D$ and ex-ante price p follows from the interaction of generating firms and final consumers on an ex-ante market, generating firms receive τ_j for each unit delivered to consumers from its pledge Δ_j, and engage in an balancing market that pays σ_j for providing capacity and $\tau_{j,im}$ for delivered imbalance energy. The resulting ex-ante price and demand is efficient when generating firms do not pay for undersupply and rationing cost, generating firms obtain payments of $\sigma_s = \tau_s = \tau_{s,im} = 0$ as well as $\left(\sigma_j, \tau_j, \tau_{j,im}\right) = \left(b_j, c_j - p, c_j\right)$ for $j = r, g$ for their engagement in the ex-ante and balancing markets, and curtailed consumers get taxed at δ for each unit of lost load.*

The advantage of a balancing market with capacity payments is that renewables firms always decide for efficient capacities (and not only for the case of perfect correlation). However, production decisions partly need to get enforced again by the system operator.

6.3 Intermediary Retailer

Suppose again the market structure where consumers pay real-time prices from Subsection 6.1 but additionally assume that an intermediary firm, that is, retailer, interacts between generating firms and final consumers. Ahead of actual consumption (ex-ante), a representative retailer R sells a fixed amount of demand Δ_R at the retail price p to final consumers (that are not willing to pay real-time prices) and buys generation y_R either from the *day-ahead* or *real-time market* at price ρ (17). We consider perfectly competitive retailers and thus assume that there is a continuum of i retailers that can decide to enter the market, so that $\int_i \Delta_R(i)\,di + \sum_j \int_n \Delta_j(n)\,dn = D$ and $\int_i y_R(i)\,di + \sum_j \int_n y_j(n)\,dn = Y$. Retailers do not account for their influence on prices and events that might materialize (as generating firms).

In the events of renewables curtailed (Ω_1) and gas power dispatched (Ω_2), the electricity produced is sufficient to serve the whole demand so that $\Delta_R = y_R$ (and $D = Y$). In case of lost load (Ω_3), some retailers cannot buy enough electricity so that $y_R < \Delta_R$. The consumers with the lowest willingness-to-pay need to get curtailed (at the burden of rationing cost δ) and compensated at $c_L$from the system operator. The retailer carries rationing cost and pays for undersupply. Moreover, it does not matter whether or not the retailer buys day-ahead or in real-time because final prices are the same. Expected profits are

$$\begin{aligned} E[\pi_R] &= p\left(\Delta_R \Pr_{12} + E[y_R|\Omega_3]\Pr_3\right) - \left(c_r\Delta_R\Pr_1 + c_g\Delta_R\Pr_2 + (c_{voll}+\delta)E[y_R|\Omega_3]\Pr_3\right) \\ &\quad -(\tau_P+\delta)E[\Delta_R - y_R|\Omega_3]\Pr_3. \end{aligned} \tag{45}$$

Retailers earn money from selling Δ_R or y_R, respectively (first term in the first line). The second term in the first line documents cost from buying electricity on the real-time market, and the second line covers cost from undersupply and the burden of rationing cost. Profit maximization (via differentiation of $E[\pi_R]$ w.r.t. Δ_R) yields the first-order condition

$$\frac{\partial E[\pi_R]}{\partial \Delta_R} = p\Pr_{12} - c_r\Pr_1 - c_g\Pr_2 - (\tau_P+\delta)\Pr_3 \le 0\ [= 0 \text{ if } \Delta_R > 0]. \tag{46}$$

Using the marginal cost of curtailment $c_{voll}+\delta$ and a compensation of $\tau_P = c_L$ (see Proposition 5), the resulting retail price is equivalent to the (efficient) ex-ante price as specified in Section 5.1, i.e., $p = c_r\Pr_1 + c_g\Pr_2 + (c_{voll}+\delta)\Pr_3 = p^*$. Profits of a retailer simplify to

$$E[\pi_R] = \Delta_R\left(p - c_r\Pr_1 - c_g\Pr_2 - (c_{voll}+\delta)\Pr_3\right). \tag{47}$$

It is straightforward to show that the zero-profit condition for retailers is fulfilled. We summarize:

Proposition 12. *Suppose that $Q_r^*, Q_g^* > 0$, intermediary firms sign contracts with final consumers to deliver Δ_R at the retail price p, and buy electricity at the real-time price ρ as specified by (17) either on a day-ahead or real-time market. The resulting retail price and demand are efficient when intermediary firms pay c_L for each unit of undersupply, carry the burden of rationing cost δ, and curtailed consumers receive c_L for each unit of lost load.*

Observe that the combination of day-ahead demand and real-time pricing with retailers is equivalent to a combination of ex-ante and real-time markets. Also combinations of both situations lead to the same outcome. Risk-neutral generating firms might supply some final consumers. Risk-neutral consumers might be willing to pay real-time prices. Retailers can then substitute for risk-averse generating firms or consumers, respectively, supply remaining consumers, and thus complete the market. However, a system operator would still be necessary to collect undersupply penalties and compensate consumers accordingly. A system operator would be obsolete when and only when there is just one retailer on the market.

6.4 Multiple Ex-ante Markets

Real-world power markets apply multiple ex-ante markets to reflect, for example, risk-aversion of firms, consumer heterogeneity, or declining levels of uncertainty. We create a situation that justifies the existence of multiple ex-ante markets by supposing that there are three types of consumers. Final consumers with willingness-to-pay (WTP) of $m = high, mid, low$ demand for D_m, obtain utility U_m, and suffer disruption cost of $c_{L,m}$. Moreover, define $c_{voll,m} := c_{L,m} + U_m^{'}(D_m)$ as the value of lost of consumer group m with $c_{voll,high} > c_{voll,mid} > c_{voll,low}$. Final consumers with a higher WTP and higher value of lost load would pay more for electricity but would receive priority dispatch in case of lost load. Consequently, we need to expand lost load events to those events where *low* WTP consumers are curtailed (denoted by $\Omega_{3,low}$), *mid* WTP consumers ($\Omega_{3,mid}$), and *high* WTP consumers ($\Omega_{3,high}$). We summarize the social planner solution in Lemma 4 (see Appendix G for proof).

Lemma 4. *Suppose that final consumers differ in utility and disruption cost so that $c_{voll,high} > c_{voll,mid} > c_{voll,low}$. Efficient demand of consumer group $m = high, mid, low$ follows from*

$$U_m^{'}(D_m) = c_r \Pr{}_1 + c_g \Pr{}_2 + \begin{cases} \left(c_{voll,low}+\delta\right)\left(\Pr{}_{3,low}+\Pr{}_{3,mid}+\Pr{}_{3,high}\right) & m = low, \\ \left(c_{voll,low}+\delta\right)\Pr{}_{3,low}+\left(c_{voll,mid}+\delta\right)\left(\Pr{}_{3,mid}+\Pr{}_{3,fwm}\right) & m = mid, \\ \left(c_{voll,low}+\delta\right)\Pr{}_{3,low}+\left(c_{voll,mid}+\delta\right)\Pr{}_{3,mid}+\left(c_{voll,high}+\delta\right)\Pr{}_{3,high} & m = high. \end{cases}$$

Final consumers with a low WTP are not curtailed in Ω_1, Ω_2 and respective cost are c_r, c_g. As soon as lost load occurs, low WTP consumers are curtailed first and suffer $c_{voll,low}$, whereas δ reflects additional cost from rationing. The situation slightly changes for mid WTP consumers. Those are not curtailed in $\Omega_{3,low}$ but for $\Omega_{3,mid}$ and $\Omega_{3,high}$ (and suffer $c_{voll,mid}$, additional rationing cost apply as well). Finally, high WTP consumers are only curtailed in $\Omega_{3,high}$ events where they suffer $c_{voll,high}$. However, each additional demand from the high WTP group increases *suffering* of low and mid consumers, reflected by $\left(c_{voll,low}+\delta\right)\Pr{}_{3,low}$ or $\left(c_{voll,mid}+\delta\right)\Pr{}_{3,mid}$, respectively.

We can now transfer those results to a firm equilibrium. $\Delta_{j,eam}$ is electricity sold on an ex-ante market *eam*. We consider three ex-ante markets $eam = fwd, dam, idm$: forward markets for high WTP consumers (denoted by subscript fwd), day-ahead markets for mid WTP consumers (subscript *dam*), and intraday markets for low WTP consumers (*idm*). However, we can easily expand the model by multiple forward and intraday markets reflecting (almost) continuous trading. Firms can balance their ex-ante pledge $\Delta_{j,eam}$ by purchasing or selling $y_{j,rtm}$ ($y_{j,rtm} < 0$ reflects purchases), respectively, at the real-time price ρ (17) in a real-time market (denoted by subscript *rtm*). Proposition 13 shows that a firm equilibrium leads to the efficient outcome.

Proposition 13. *Suppose that $Q_r^*, Q_g^* > 0$, final consumers differ in utility and disruption cost so that $c_{voll,high} > c_{voll,mid} > c_{voll,low}$, and electricity is traded in forward, day-ahead, intraday, and real-time markets so that $D_{fwm} = D_{high}$, $D_{dam} = D_{mid}$, and $D_{idm} = D_{low}$. The resulting ex-ante prices p_{eam} and demand $D = \sum_{eam} D_{eam}$ are efficient when generating firms pay $\tau_{P,fwm} = c_{L,high}$, in forward markets, $\tau_{P,dam} = c_{L,mid}$ in day-ahead markets, and $\tau_{P,idm} = c_{L,low}$ in intraday markets for each unit of undersupply, generating firms carry the burden of rationing cost δ, and curtailed consumers from group m receive $c_{L,m}$ for each unit of lost load.*

The final outcome with regard to multiple ex-ante markets shows how real-world power markets already address the inflexible consumer demand, limited dispatchability of steam power plants, and uncertain supply of renewables. However, the overarching difference (to obtain an efficient outcome) is the missing consumer compensation in case of lost load and too low undersupply penalties. Firms are then forced to buy (either in real-time or ex-ante) or provide sufficient back-up capacities which in turn enhances long-run resource adequacy (and thus address the reliability externality).

7 Concluding Remarks

Designing power markets is complicated, and often done suboptimal due to many possible and even overlapping market failures. Economic issues (e.g., market power, security of service as a public good, environmental externalities), behavior (e.g., asymmetric information, risk-aversion, demand response), institutional design (e.g., contractual fixed consumer prices), and technological limitations (e.g., ramping constraints, generator flexibility, spatial distribution, stochastic supply) cause these failures. We shed new light on these issues by addressing the role of *ramping constraints* (e.g., coal and nuclear power plants cannot react to changing supply in real-time) and *generator flexibility* (e.g., gas power can react instantly), *spatial distribution* and *stochastic supply* of intermittent renewable energies, the *lack of demand response* and possible *curtailment of final consumers*, and how those potential failures are addressed by existing designs. We draw five main recommendations for electricity market design that tackle how electricity is priced efficiently.

First, *ramping constraints* are not the source of market failures but *generator flexibility* in combination with ex-ante pricing establishes a missing market problem. We model ramping constraints and generator flexibility by introducing a sequential dispatch decision model. Steam power production needs to be scheduled ahead of the realization of the random supply of intermittent renewables and cannot be changed later on, whereas gas power is able to react instantly. We detect that ramping constraints do not lead to inefficiencies, whereas generator flexibilities do so under *ex-ante pricing*. Firms using technologies with ramping constraints (such as steam power) provide efficient capacities and recover cost (no market failure), because those technologies are dispatched ex-ante (no output risk) and, thus, both ex-ante and real-time prices reflect the value of electricity produced. Firms using flexible generators (such as gas power) cannot recover cost and provide no capacity at all (market failure) under ex-ante pricing, because the adjustment of production in response to fluctuating supply of renewables (output risk) is inadequately priced. Additionally, the existence of a cheaper conventional technology with ramping constraints (i.e., steam power) keeps prices below the necessary threshold for cost recovery. Support mechanisms resolve the failure, but capacity payments alone are never sufficient to do so. One might argue that the underlying market failure is caused by the technologies with ramping constraints, but it is indeed a pricing issue. Real-time pricing addresses the problem by pricing the possible curtailment of consumers. Now, gas power firms recover cost and provide sufficient capacities for resource adequacy. Support mechanisms can overcome the problem of pricing generator flexibility but the real-time market solution is preferable due to complicated tax-subsidy-schemes when deciding for ex-ante pricing.

Second, renewables such as wind and solar power are subject to an output risk when renewables generators do not face the same weather conditions. We model *spatial distribution* and *stochastic supply* of intermittent renewables by introducing correlation of single units with total uncertain renewables supply. Even real-time pricing does not provide efficient long-run incentives for renewables firms due to the spatial distribution of intermittent renewables. The availability of wind turbines and solar PV panels are weather-dependent. Weather conditions are spatially correlated, but never perfectly, which would be necessary to arrive at the efficient outcome. We focus our analysis on situations where at least two technologies—gas power and renewables—are present in equilibrium. Here, the magnitude of the output risk increases with growing shares of renewables. Consequently, spatially small markets (higher correlation) and those with few renewables capacities does not suffer much from that market failure, whereas spatially bigger markets and those with higher shares of renewables generation need to address that market failure appropriately. As a consequence, renewables technologies either require additional subsidization to overcome the impact of the output risk on firm's revenues or price zones need to be flexibly clustered to zones with same (wind and solar) weather conditions. Such *weather clustering zonal pricing* in combination with locational marginal pricing (LMP) might also steer renewables investment efficiently into the right locations.

Third, curtailed consumers need to receive compensation in high of disruption cost when they pay ex-ante prices and need to get taxed in high of rationing cost when they pay real-time prices. The sequential decision structure of the model also reflects the *lack of demand response* and possible *curtailment of final consumers*. Socially optimized systems balance the possibility of a shortage of supply and the holding of (often unused) back-up capacity. As back-up capacity does, consumers offer system flexibility in means of curtailment and, thus, need appropriate compensation for curtailment. When consumers pay ex-ante prices a subsidy in the height of disruption cost compensates curtailed consumers and turn them indifferent to getting curtailed or not. This result underlines the importance of a compensation mechanism for curtailed consumers. Curtailed consumers that pay real-time prices need to get taxed in high of rationing costs. In lost load events, the efficient real-time price rises above the VOLL by rationing costs, so that the real-time price is higher than the opportunity cost of not consuming electricity. Additionally, non-reactive consumers do not reduce the overall burden of rationing cost, whereas back-up capacity does so. As a consequence, the planner needs to impose a tax in height of rationing cost to enforce efficient demand.[39] This result underlines that scarcity pricing at the (marginal) VOLL is inefficient when the system operators seeks to curtail—at the burden of rationing cost—only those consumers with the lowest marginal utility. Rationing cost should be reflected by prices in those rare lost load

[39] Remember that curtailment of consumers is socially optimal as soon as intermittent renewables (that are subject to stochastic supply) are competitive. It is a mixture of engineering and behavioral studies to determine the VOLL (e.g., Willis and Garrod, 1997; JRC, 2018) and a following economic optimization problem (e.g., Ovaere et al., 2019) to find the socially optimal probability of curtailment or reliability level, respectively.

events, opening the case for taxing curtailed consumers that pay real-time prices and intensifying the discussion for including cost of (transmission and distribution) system operators in wholesale market prices. However, final consumers prefer paying ex-ante prices because they do not respond in real time and are generally risk-avers. Risk-neutral generating firms can sell ex-ante (over-the-counter, in forward, day-ahead, or intraday markets) to final consumers or retailers can engage as intermediary firm between (risk-avers) generating firms and final consumers to complete markets.

Fourth, there are *three missing pieces for efficiency pricing* in Northern American (with day-ahead and real-time markets) as well as European markets designs (that combine ex-ante with balancing markets). First, undersupply penalties when firms are not able to fulfill their ex-ante (forward, day-ahead, or intraday) delivery pledge either by own generation or trading must be set equal to disruption cost of final consumers. Those disruption cost are the dominating part of the VOLL (the other one are marginal utility losses) and thus penalties might be in the range of 5,000 to 10,000 €/MWh. Next, consumers generally pay ex-ante prices and thus the consumer compensation in lost load events is missing. Such a compensation for bad ex-ante delivery pledges is not implemented yet in real-world power markets. The system operator would serve as clearing entity that receives payments from generating firms for undersupply and transfers those payments to the curtailed consumers. The role of the system operator is crucial here because generating firms or intermediaries such as retailers cannot transfer undersupply penalties directly to consumers, because only those consumers with the lowest WTP get curtailed. Thus, the system operator matches firms' undersupply with curtailment of consumers. Moreover, consumer curtailment can also be reinterpreted as consultation of some additional (operating) reserves that runs at the same cost as final consumers would be curtailed otherwise. Lost load events would then be reserve events. Compensation (or price) for reserves is widespread implemented. Retailers or generating firms with contracts to deliver final consumers would automatically back-up their delivery pledges by buying on such an additional market. Finally, flexible pricing zones in accordance with weather conditions resolves the output risk stemming from spatial distribution and stochastic supply of intermittent renewable energies.

Fifth, capacity payments or markets (besides already existing ones to safeguard the public good *security of services*) could be justified either by capped real-time prices (in Northern American markets) or by balancing markets with capacity payments (in European markets). Price caps reduce the scarcity price in lost load events below the marginal cost of curtailment (VOLL plus rationing cost). Such capping segregates markets by technologies. Steam power and renewables firms would engage on a day-ahead market and pay undersupply penalties in high of disruption cost for each unit of undersupply. Gas power firms in turn would cover the entire real-time market alone and receive capacity payments that close the missing money from capping prices. Efficient balancing markets with capacity payments in turn subsidize full capacity expenses for renewables and gas power but not for steam power, whereas marginal profits from selling electricity are deprived.

There are a few limitations in our analysis. For example, we consider just the two extreme cases of correlation. Following Chao (1983), one might implement a correlation measure—as he does for demand—on the supply side. Note that this would not change our main finding that a real-time price is not sufficient for renewables firms to recover cost, because they face an output risk.[40] The main restriction is the assumption of very strict dispatchability levels. For example, support mechanisms under ex-ante pricing focus on consumers and flexible generators. In reality, steam power is at least partially able to react instantly to stochastic fluctuations in the supply of renewables, albeit at a higher cost. Indeed, modern steam power has higher ramp-up and ramp-down possibilities and lower cost than was formerly the case. This partial flexibility opens the case for participation of nuclear and coal power plants in capacity mechanisms. The analysis further neglects the potential impact of storage to balance stochastic supply of renewable energies.[41] Investigating the effect of flexible steam power in a dynamic investment setup with inflexible steam power, renewables, gas power, and storage options would be an interesting and useful topic for future work.

Finally, classic peak-load pricing (see Crew et al., 1995) suggest technology composites of baseload (mainly steam power) and peakload power plants (mainly gas power) to balance demand fluctuations. We neglect periodic demand in our basic model and, thus, some of our findings refer only to gas power dedicated to balance renewables output. However, our findings still allow to draw recommendations for real-world power markets with periodic demand, because we show that the main findings, that is, markets are incomplete for renewables firms under real-time pricing and curtailed consumers demand for compensation, still hold true under periodic demand..

References

Ambec, S. and C. Crampes (2012). Electricity provision with intermittent sources of energy. *Resource and Energy Economics 34*(3), 319–336.

Antweiler, W. (2017). A two-part feed-in-tariff for intermittent electricity generation. *Energy Economics 65*, 458–470.

Bajo-Buenestado, R. (2017). Welfare implications of capacity payments in a price-capped electricity sector: A case study of the texas market (ercot). *Energy Economics 64*, 272–285.

[40] Moreover, correlation in the supply of generating units is a predominant problem for intermittent renewable energies. However, steam power plants also face such correlations. For example, nearby located steam power plants have to shut down when their cooling river gets too warm. Using a similar model, we could deduct that even dispatchable generators need additional payments under real-time pricing.

[41] Recent studies (e.g., Helm and Mier, 2021; Schmalensee, 2019) show—in models that neglect sequential dispatch decisions and different pricing schemes—that storage does not induce inefficiencies.

Bejan, I., C. L. Jensen, L. M. Andersen, and L. G. Hansen (2019). The hidden cost of real time electricity pricing. *IFRO Working Paper No. 2019/03*.

Boiteux, M. (1949). La tarification des demandes en point: application de la theorie de la vente au cout marginal. *Revue Generale de l'Electicite 58, 321–340, 1960 translated as "Peak Load Pricing." Journal of Business 33(2), 157–179*.

Borenstein, S. and J. B. Bushnell (2018). Do two electricity pricing wrongs make a right? Cost recovery, externalities, and efficiency. *NBER Working Paper No. 24756*.

Brown, G. and M. B. Johnson (1969). Public utility pricing and output under risk. *American Economic Review 59*(1), 119–128.

Bublitz, A., D. Keles, F. Zimmermann, C. Fraunholz, and W. Fichtner (2019). A survey on electricity market design: Insights from theory and real-world implementations of capacity remuneration mechanisms. *Energy Economics 80*, 1059–1078.

Bulow, J. and P. Klemperer (2012). Regulated prices, rent seeking, and consumer surplus. *Journal of Political Economy 120*(1), 160–186.

Bushnell, J., E. T. Mansur, and K. Novan (2017). Review of the economics literature on us electricity restructuring. *Unpublished manuscript, Department of Economics, University of California at Davis*.

Bye, R. T. (1926). The nature and fundamental elements of costs. *The Quarterly Journal of Economics*, 30–62.

Bye, R. T. (1929). Composite demand and joint supply in relation to public utility rates. *The Quarterly Journal of Economics*, 40–62.

Carlton, D. W. (1977). Peak load pricing with stochastic demand. *American Economic Review 67*(5), 1006–1010.

Chao, H.-P. (1983). Peak load pricing and capacity planning with demand and supply uncertainty. *The Bell Journal of Economics 14*(1), 179–190.

Chao, H.-P. (2011). Efficient pricing and investment in electricity markets with intermittent resources. *Energy Policy 39*(7), 3945–3953.

Costello, K. (2012). Should utilities compensate customers for service interruptions? *The Electricity Journal 25*(7), 45–55.

Crampes, C. and J. Renault (2019). How many markets for wholesale electricity when supply is partially flexible? *Energy Economics 81*, 465–478.

Crampes, C. and D. Salant (2018). A multi-regional model of electric resource adequacy. *TSE Working Paper, n. 17-877*.

Cramton, P. (2017). Electricity market design. *Oxford Review of Economic Policy 33*(4).

Cramton, P., A. Ockenfels, and S. Stoft (2013). Capacity market fundamentals. *Economics of Energy & Environmental Policy 2*(2), 27–46.

Crew, M. A., C. S. Fernando, and P. R. Kleindorfer (1995). The theory of peak-load pricing: A survey. *Journal of Regulatory Economics 8*(3), 215–248.

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

EC (2016). Metis technical note t4: Overview of european electricity markets. Technical report, European Commission.

Eisenack, K. and M. Mier (2019). Peak-load pricing with different types of dispatchability. *Journal of Regulatory Economics 56*(2-3), 105–124.

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

Fabra, N., N.-H. M. Von der Fehr, and M.-Á. De Frutos (2011). Market design and investment incentives. *The Economic Journal 121*(557), 1340–1360.

Green, R. J. and T.-O. Léautier (2017). Do costs fall faster than revenues? Dynamics of renewables entry into electricity markets. *TSE Working Paper, n. 15-591, revised version*.

Helm, C. and M. Mier (2019). On the efficient market diffusion of intermittent renewable energies. *Energy Economics 80*, 812–830.

Helm, C. and M. Mier (2021). Steering the energy transition in a world of intermittent electricity supply: Optimal subsidies and taxes for renewables and storage. *Journal of Environmental Economics and Management*, 102497.

Hirshleifer, J. (1958). Peak loads and efficient pricing: Comment. *The Quarterly Journal of Economics 72*(3), 451–462.

Hogan, W. W. (2005). On an “energy only” electricity market design for resource adequacy.

Houthakker, H. S. (1951). Electricity tariffs in theory and practice. *The Economic Journal 61*(241), 1–25.

Joskow, P. and J. Tirole (2006). Retail electricity competition. *The Rand Journal of Economics 37*(4), 799–815.

Joskow, P. and J. Tirole (2007). Reliability and competitive electricity markets. *The Rand Journal of Economics 38*(1), 60–84.

Joskow, P. L. (2011). Comparing the costs of intermittent and dispatchable electricity generating technologies. *American Economic Review 101*(3), 238–241.

Joskow, P. L. (2019). Challenges for wholesale electricity markets with intermittent renewable generation at scale: the us experience. *Oxford Review of Economic Policy 35*(2), 291–331.

JRC (2018). Jrc science for policy report: Societal appreciation of energy security. Technical report, Joint Research Center.

Keppler, J. H. (2017). Rationales for capacity remuneration mechanisms: Security of supply externalities and asymmetric investment incentives. *Energy Policy 105*, 562–570.

Kleindorfer, P. R. and C. S. Fernando (1993). Peak-load pricing and reliability under uncertainty. *Journal of Regulatory Economics 5*(1), 5–23.

Küfeoğlu, S. et al. (2015). Economic impacts of electric power outages and evaluation of customer interruption costs.

Kumano, T. (2011). A functional optimization based dynamic economic load dispatch considering ramping rate of thermal units output. *Power Systems Conference and Exposition*, 1–8.

Leautier, T.-O. (2018). On the long-term impact price caps: Investment, uncertainty, imperfect competition, and rationing. *International Journal of Industrial Organization 61*, 53–95.

Lesser, J. A. and X. Su (2008). Design of an economically efficient feed-in tariff structure for renewable energy development. *Energy Policy 36*(3), 981–990.

Milstein, I. and A. Tishler (2019). On the effects of capacity payments in competitive electricity markets: Capacity adequacy, price cap, and reliability. *Energy Policy 129*, 370–385.

Neuhoff, K. and L. De Vries (2004). Insufficient incentives for investment in electricity generations. *Utilities Policy 12*(4), 253–267.

Newbery, D. (2016). Missing money and missing markets: Reliability, capacity auctions and interconnectors. *Energy Policy 94*, 401–410.

Newbery, D., M. G. Pollitt, R. A. Ritz, and W. Strielkowski (2018). Market design for a high-renewables european electricity system. *Renewable and Sustainable Energy Reviews 91*, 695–707.

Novan, K. (2015). Valuing the Wind: Renewable Energy Policies and Air Pollution Avoided. *American Economic Journal: Economic Policy 7*(3), 291–326.

Oren, S. S. and J. A. Doucet (1990). Interruption insurance for generation and distribution of electric power. *Journal of Regulatory Economics 2*(1), 5–19.

Ovaere, M., E. Heylen, S. Proost, G. Deconinck, and D. Van Hertem (2019). How detailed value of lost load data impact power system reliability decisions. *Energy Policy 132*, 1064–1075.

Paris Agreement, P. (2015). Paris agreement. In *Report of the Conference of the Parties to the United Nations Framework Convention on Climate Change (21st Session, 2015: Paris)*.

Pinho, J., J. Resende, and I. Soares (2018). Capacity investment in electricity markets under supply and demand uncertainty. *Energy 150*, 1006–1017.

Schill, W.-P., M. Pahle, and C. Gambardella (2017). Start-up costs of thermal power plants in markets with increasing shares of variable renewable generation. *Nature Energy 2*(17050), 1–6.

Schmalensee, R. (2019). On the efficiency of competitive energy storage. *Available at SSRN*.

Schwenen, S. (2014). Market design and supply security in imperfect power markets. *Energy Economics 43*, 256–263.

Steiner, P. O. (1957). Peak loads and efficient pricing. *The Quarterly Journal of Economics 71*(4), 585–610.

Sunar, N. and J. R. Birge (2019). Strategic commitment to a production schedule with uncertain supply and demand: Renewable energy in day-ahead electricity markets. *Management Science 65*(2), 714–734.

Turvey, R. (1968). Peak-load pricing. *Journal of Political Economy 76*(1), 101–113.

Turvey, R. and D. Anderson (1977). *Electricity economics: Essays and case studies*. Johns Hopkins University Press, Baltimore.

Visscher, M. L. (1973). Welfare-maximizing price and output with stochastic demand: Comment. *American Economic Review 63*(1), 224–229.

Wallnerström, C. J. (2008). *On risk management of electrical distribution systems and the impact of regulations*. Ph. D. thesis, KTH.

Williamson, O. E. (1966). Peak-load pricing and optimal capacity under indivisibility constraints. *American Economic Review 56*(4), 810–827.

Willis, K. and G. Garrod (1997). Electricity supply reliability: Estimating the value of lost load. *Energy Policy 25*(1), 97–103.

Wilson, R. (2002). Architecture of power markets. *Econometrica 70*(4), 1299–1340.

Wolak, F. A., J. Glanchant, P. Joskow, and M. Pollitt (2020). Wholesale electricity market design. *Handbook on the Economics of Electricity Markets 18*.

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

Appendix

A Proof of Lemma 1

We can solve the problem by using Kuhn-Tucker conditions. To keep it brief, we do it in more intuitive way. We know that lost load occurs when production is insufficient to meet load. Thus, lost load is given by

$$D_l = \begin{cases} D - Y_s - Y_g - Y_r & \text{for } Y_s + Y_r + Y_g < D, \\ 0 & \text{else.} \end{cases} \tag{48}$$

In Stage 5, D, Y_s, Y_r are given. We can leave out expectations and maximize welfare w.r.t. Y_g. We obtain

$$\frac{\partial W}{\partial Y_g} = \begin{cases} -c_g + c_l + \delta + U^{'}(Y) & > 0 \text{ for } Y_s + Y_r + Y_g < D, \\ -c_g & < 0 \text{ else.} \end{cases} \tag{49}$$

Note that marginal utility in (49) is evaluated at aggregate production $Y = Y_s + Y_r + Y_g$. Gas power produces to avoid lost load. The optimal production schedule is

$$Y_g = \begin{cases} Q_g & \text{for } Y_s + Y_r + Y_g < D, \\ D - Y_s - Y_r & \text{for } Y_s + Y_r < D < Y_s + Y_r + Y_g, \\ 0 & \text{else.} \end{cases} \tag{50}$$

For renewables production, we know that lost load occurs if $Y_r < D - Y_s - Q_g$, where we anticipated that $Y_g = Q_g$. Moreover, D, Y_s are given at Stage 4 so that we can leave out expectations again and maximize welfare w.r.t. Y_r to obtain

$$\frac{\partial W}{\partial Y_r} = \begin{cases} -c_r + c_g + c_u + U^{'}(Y) & > 0 \text{ for } Y_r < D - Y_s - Q_g, \\ -c_r + c_g & > 0 \text{ for } D - Y_s - Q_g \le Y_r < D - Y_s, \\ -c_r & < 0 \text{ else.} \end{cases} \tag{51}$$

Renewables produce to avoid gas turbine production. The optimal production schedule is

$$Y_r = \begin{cases} \tilde{Q}_r & \text{for } \tilde{Q}_r < D - Y_s - Y_g, \\ \tilde{Q}_r & \text{for } X - Y_s - Y_g \leq \tilde{Q}_r < D - Y_s, \\ D - Y_s & \text{else.} \end{cases} \quad (52)$$

From (50) and (52), we obtain the interval of events $\Omega_1 = [D - Q_s, Q_r]$, $\Omega_2 = [D - Q_s - Q_g, D - Q_s)$, and $\Omega_3 = [0, D - Q_s - Q_g)$. For example, $Y_s + Y_r + Y_g < D$ is equivalent to Ω_3. Using these intervals and (52), (50), and (48), we obtain:

$$Y_r^* = \begin{cases} \tilde{Q}_r & \text{for } \tilde{Q}_r \in \Omega_3, \\ D - Q_s & \text{else,} \end{cases} \quad (53)$$

$$Y_g^* = \begin{cases} Q_g & \text{for } \tilde{Q}_r \in \Omega_3, \\ D - Q_s - \tilde{Q}_r & \text{for } \tilde{Q}_r \in \Omega_2, \\ 0 & \text{else,} \end{cases} \quad (54)$$

$$D_l = \begin{cases} D - Q_s - Q_g - Q_r & \text{for } \tilde{Q}_r \in \Omega_3, \\ 0 & \text{else.} \end{cases} \quad (55)$$

Note that D_l is not a decision and thus cannot be optimized (no asterisk applies here).

Turning to Stage 3 and using the fact that $\tilde{Q}_r$ is boundedly integrable, we can interchange differentiation and expectation (see Chao, 1983, 2011; Eisenack and Mier, 2019). Using (53) to (55), we differentiate (1) conditional to the realized interval of events. Taking expectations yields

$$\frac{\partial E[W]}{\partial Y_s} = c_r \Pr{}_1 + c_g \Pr{}_2 + \left(c_L + \delta + U'(Y)\right) \Pr{}_3 - c_s. \quad (56)$$

Suppose that $\frac{\partial E[W]}{\partial Y_s} < 0$ so that $Y_s < Q_s$. Then, in Stage 1, maximization of welfare w.r.t. Q_s (we need to apply the same steps as for (56)) yields $\frac{\partial E[W]}{\partial Q_s} = -b_s < 0$. It follows that $Y_s = Q_s = 0$, a contradiction to $Y_s < Q_s$. We conclude that $\frac{\partial E[W]}{\partial Y_s} > 0$ so that $Y_s = Q_s$.

B Proof of Lemma 2

From (55), we know that $X = Y$ for $\tilde{Q}_r \in \Omega_3$ and $X = D$ for $\tilde{Q}_r \in 12$. We obtain:

$$U(X) = \begin{cases} U(Y) & \text{for } \tilde{Q}_r \in \Omega_3, \\ U(D) & \text{else}, \end{cases} \tag{57}$$

where $Y = Q_s + \tilde{Q}_r + Q_g$. As in Appendix A, we use the fact that $\tilde{Q}_r$ is boundedly integrable and, thus, interchange the order of differentiation and expectations to maximize welfare with regard to demand. We differentiate (1) conditional to the respective interval Ω_c, thereby using $Y_s = Q_s$ and (53) to (55),

$$\frac{\partial W}{\partial D} = \begin{cases} U'(D) - c_r & \text{for } \tilde{Q}_r \in \Omega_1, \\ U'(D) - c_g & \text{for } \tilde{Q}_r \in \Omega_2, \\ -(c_L + \delta) & \text{else}. \end{cases}$$

Taking expectations yields the first-order condition:

$$\frac{\partial E[W]}{\partial D} = U'(D)\,\mathrm{Pr}_{12} - c_r \mathrm{Pr}_1 - c_g \mathrm{Pr}_2 - (c_L + \delta)\,\mathrm{Pr}_3 \leq 0\,[= 0 \text{ if } D^* > 0]. \tag{58}$$

In the model setup, we assumed that the Inada conditions are fulfilled, that is, the first-order condition of demand (58) must bind for at least some $D > 0$. Using $U'(D)\,\mathrm{Pr}_{12} = U'(D) - U'(D)\,\mathrm{Pr}_3$ to solve the binding condition for U', enables us to find the marginal utility to maximize welfare with regard to demand as shown in Lemma 2.

C Proof of Lemma 3

Start with $Q_g^* > 0$. From (4) and (5), we obtain:

$$\mathrm{Pr}_3^* = \frac{b_g}{c_{voll} + \delta - c_g}, \tag{59}$$

$$\mathrm{Pr}_2^* = \frac{b_r - (c_{voll} + \delta - c_r)\,\overline{a}_3\,\mathrm{Pr}_3^*}{(c_g - c_r)\,\overline{a}_2}. \tag{60}$$

Next, suppose that $Q_g^* = 0$. (4) is not binding and thus $\mathrm{Pr}_2^* = 0$. From (5), we obtain:

$$\Pr_3^* = \frac{b_r}{(c_{voll}+\delta-c_r)\,\overline{a}_3}. \tag{61}$$

D Proof of Proposition 1

Suppose that $q_r > 0$, $Q_g^* = 0 \Leftrightarrow \Pr_2^* = 0$, and $U'(D^*) \leq b_s + c_s$. The planner enforces $y_r < \tilde{q}_r$ so that $Y_r = D - Q_s < \tilde{Q}_r$ in Ω_1 and $y_r = \tilde{q}_r$ so that in Ω_3. Expected production is $E[y_r] = E[\tilde{q}_r|\Omega_3]\Pr_3 + E[y_r|\Omega_1]\Pr_1$. From (8), we obtain $\frac{\partial E[\pi_r]}{\partial q_r} = (p-c_r)\,\overline{a}_3 \Pr_3 - b_r = 0$. Substituting $p = c_r + (c_v - c_r)\Pr_3$ (by using $c_r \Pr_1 = c_r - c_r \Pr_3$) yields

$$\begin{aligned}
0 &= (c_r + (c_{voll}+\delta-c_r)\Pr_3 - c_r)\,\overline{a}_3 \Pr_3 - b_r q_r \\
&= (c_{voll}+\delta-c_r)\,\overline{a}_3 (\Pr_3)^2 - b_r.
\end{aligned}$$

Noting that $\Pr_3$ cannot be negative, we can solve this for $\Pr_3 = \sqrt{\Pr_3^*}$ (see Lemma 3). Note that $p \leq b_s + c_s$ due to the existence of steam power. $\Pr_3 = 1$ would lead to $p = c_{voll} + \delta > b_g + c_g > b_s + c_s$ and violate $p \leq b_s + c_s$. It follows that $1 > \Pr_3 = \sqrt{\frac{b_r}{(c_{voll}+\delta-c_r)\overline{a}_3}} > \Pr_3^*$. The equilibrium probability of lost load is higher than the efficient probability, because renewable capacity is below its efficient level. This also yields to higher prices since $p = c_r \Pr_1 + (c_{voll}+\delta)\Pr_3 = c_r + (c_{voll}+\delta-c_r)\Pr_3 > c_r + (c_{voll}+\delta-c_r)\Pr_3^*$.

Turn to profits. Substituting p and do some rearrangering, yields

$$\begin{aligned}
E[\pi_r] &= (c_r + (c_{voll}+\delta-c_r)\Pr_3 - c_r)\,E[y_r] - b_r q_r \\
&= (c_{voll}+\delta-c_r)\left(\Pr_3 E[y_r] - \frac{b_r}{(c_{voll}+\delta-c_r)\,\overline{a}_3}\overline{a}_3 q_r\right) \\
&= (c_{voll}+\delta-c_r)\left(\Pr_3 E[y_r] - (\Pr_3)^2\,\overline{a}_3 q_r\right).
\end{aligned}$$

For *corr*, we know that $E[y_r] = E[\tilde{q}_r|\Omega_3]\Pr_3 + E[y_r|\Omega_1]\Pr_1$ and $a_3 q_r \Pr_3 = E[\tilde{q}_r|\Omega_3]\Pr_3$. It follows that $E[\pi_r] = (c_v - c_r)\Pr_3 E[y_r|\Omega_1]\Pr_1 > 0$. Note that production of firms is perfectly correlated and all firms produce the same for each event. Thus, zero profits only occur if $\Pr_1 = E[y_r|\Omega_1] = 0$ and $\Pr_3 = 1$, which violates $p \leq b_s + c_s$ as argued above. We conclude that $E[y_r|\Omega_1] > 0$ and profits are positive.

For *ind*, we have $\overline{a}_3 q_r = a q_r = E[\tilde{q}_r]$ and it follows that $E[\pi_r] = (c_{voll}+\delta-c_r)\Pr_3 (E[y_r] - E[\tilde{q}_r]\Pr_3)$.

E Calculations for Table 1

We now demonstrate how to calculate $E[\pi_r] = -a_2 q_r \gamma$. Suppose that $\tau_1 = c_r - p$, $\tau_2 = c_g - p$, and $\tau_3 = c_{voll} + \delta - p$. From (16), profits of renewables firms are

$$E[\pi_r] \quad = \quad (c_g - c_r) E[\tilde{q}_r|\Omega_2] \Pr{}_2 + (c_{voll} + \delta - c_r) E[\tilde{q}_r|\Omega_3] \Pr{}_3 - (b_r - \sigma_r) q_r. \tag{62}$$

Note that in Ω_1 marginal production profits are zero due to $\tau_1 = c_r - p$. The efficiency condition in (14) becomes

$$b_r - \sigma_r \quad = \quad (c_g - c_r) a \Pr{}_2 + (c_{voll} + \delta - c_r) a \Pr{}_3,$$

where we have used that $\overline{a}_c = a$ for *ind*. Substituting marginal capacity cost after transfers in (62), we obtain expected profits of

$$E[\pi_r] \quad = \quad -[(a - a_2)(c_g - c_r) \Pr{}_2 + (a - a_3)(c_{voll} + \delta - c_r) \Pr{}_3] q_r \tag{63}$$

where we have used that $E[\tilde{q}_r|\Omega_2] = a_2 q_r$ and $E[\tilde{q}_r|\Omega_3] = a_3 q_r$. We use (60) and—after some rearranging—obtain $E[\pi_r] = -a_2 q_r \gamma$.

We now turn to the third line in Table 1. Accounting for $\tau^+_{r,3}$, expected profits change from (62) to

$$E[\pi_r] \quad = \quad (c_g - c_r) E[\tilde{q}_r|\Omega_2] \Pr{}_2 + \left(c_{voll} + \delta - c_r + \tau^+_{r,3}\right) E[\tilde{q}_r|\Omega_3] \Pr{}_3 - (b_r - \sigma_r) q_r.$$

The new efficiency condition becomes

$$b_r - \sigma_r \quad = \quad (c_g - c_r) a \Pr{}_2 + \left(c_{voll} + \delta - c_r + \tau^+_{r,3}\right) a \Pr{}_3 \tag{64}$$

and new expected profits are

$$\begin{aligned} E[\pi_r] \quad &= \quad -(c_g - c_r)(a - a_2) q_r \Pr{}_2 - \left(c_{voll} + \delta - c_r + \tau^+_{r,3}\right)(a - a_3) q_r \Pr{}_3 \\ &= \quad -a_2 q_r \gamma - q_r \tau^+_{r,3} (a - a_3) \Pr{}_3. \end{aligned}$$

This can be solved to obtain the value for $\tau_{r,3}^{x+}$ in Table 1. Rearranging the new efficiency condition of (64) yields

$$\begin{aligned}\sigma_r &= b_r - (c_g - c_r)\, a\,\mathrm{Pr}_2 - (c_{voll} + \delta - c_r)\, a\,\mathrm{Pr}_3 - \tau_{r,3}^{+} a\,\mathrm{Pr}_3 \\ &= \frac{aa_2}{a - a_3}\gamma.\end{aligned}$$

F Proof of Proposition 4

Subscript t denotes variables and parameters referring to period t, e.g., a_t is the average availability of renewables. Suppose that $p_t = c_r \mathrm{Pr}_{1t} + c_g \mathrm{Pr}_{2t} + c_v \mathrm{Pr}_{3t}$ is the ex-ante price in period t. Expected profits of firms are given by

$$\begin{aligned}E[\pi_s] &= \sum_t (c_r \mathrm{Pr}_{1t} + c_g \mathrm{Pr}_{2t} + c_v \mathrm{Pr}_{3t} - c_s)\, y_{st} - b_s q_s \\ E[\pi_g] &= \sum_t (c_{voll} + \delta - c_g)\, q_g \mathrm{Pr}_{3t} - b_g q_g \\ E[\pi_r] &= \sum_t (c_g - c_r)\, E\left[\tilde{q}_{jt}|\Omega_{2t}\right] \mathrm{Pr}_{2t} + \sum_t (c_{voll} + \delta - c_r)\, E\left[\tilde{q}_{jt}|\Omega_{3t}\right] \mathrm{Pr}_{3t} - b_r q_r\end{aligned}$$

Production decisions in Stages 4 and 5 are straightforward to production decisions in the one period setup. However, production decisions of steam power firms in Stage 3 requires modification. Differentiation of $E[\pi_s]$ w.r.t. y_{st} yields

$$\frac{\partial E[\pi_s]}{\partial y_{st}} = c_r \mathrm{Pr}_{1t} + c_g \mathrm{Pr}_{2t} + (c_{voll} + \delta)\, \mathrm{Pr}_{3t} - c_s.$$

The level of the first-order condition defines the subsets of periods H, L, N. For all $t \in H$, production is *highest* due to $\frac{\partial E[\pi_s]}{\partial y_{st}} > 0$ so that steam power firms use their whole capacity, $y_{st} = q_s$. If $\frac{\partial E[\pi_s]}{\partial y_{st}} = 0$ for all $t \in L$, then production is *lower*. Every production decision between zero and full capacity is optimal, $y_{st} \in (0, q_s)$. Finally, for all $t \in N$, steam power firms decide for *no* production due to $\frac{\partial E[\pi_s]}{\partial y_{st}} < 0$. Note that N and L might be empty but H never. Steam power capacity is costly and installing never used capacity cannot be optimal. Note that this specification is one crucial difference to the one-period setup. We derive the first-order conditions of firms by differentiating profits w.r.t. q_j for $j = s, g, r$:

$$
\begin{aligned}
\frac{\partial E\left[\pi_s\right]}{\partial q_s} &= \sum_{t\in H}\left(c_r \Pr_{1t} + c_g \Pr_{2t} + \left(c_{voll}+\delta\right)\Pr_{3t} - c_s\right) - b_s,\\
\frac{\partial E\left[\pi_g\right]}{\partial q_g} &= \sum_{t}\left(c_{voll}+\delta-c_g\right)\Pr_{3t} - b_g,\\
\frac{\partial E\left[\pi_r\right]}{\partial q_r} &= \sum_{t}\left(\left(c_g-c_r\right)\overline{a}_{2t}\Pr_{2t} + \left(c_{voll}+\delta-c_r\right)\overline{a}_{3t}\Pr_{3t}\right) - b_r.
\end{aligned}
$$

Remember that steam power firms do not produce in periods $t \in N$ and face zero marginal production profits for all $t \in L$. We can use this and substitute the first-order condition of steam power firms to obtain

$$
E\left[\pi_s\right] = \sum_{t\in H}\left(c_r \Pr_{1t} + c_g \Pr_{2t} + \left(c_{voll}+\delta\right)\Pr_{3t} - c_s\right) y_s - b_s q_s = 0.
$$

It is straightforward to show that gas power firms also make zero profits. We just need to substitute the first-order condition of gas power into the respective profits function to obtain $E\left[\pi_g\right] = 0$. Profits of renewables firms are more complex to determine. We can substitute the first-order condition into the profits function to obtain

$$
\begin{aligned}
E\left[\pi_r\right] &= \left[\sum_t \left(c_g-c_r\right) a_{2t}\Pr_{2t} + \sum_t \left(c_{voll}+\delta-c_r\right) a_{3t}\Pr_{3t} - b_r\right] q_r\\
&= \left[\sum_t \left(c_g-c_r\right) a_{2t}\Pr_{2t} + \sum_t \left(c_{voll}+\delta-c_r\right) a_{3t}\Pr_{3t}\right] q_r\\
&\quad -\left[\sum_t \left(c_g-c_r\right) \overline{a}_{2t}\Pr_{2t} + \sum_t \left(c_{voll}+\delta-c_r\right) \overline{a}_{3t}\Pr_{3t}\right] q_r\\
&= \left[\sum_t \left(c_g-c_r\right)\left(a_{2t}-\overline{a}_{2t}\right)\Pr_{2t} + \sum_t \left(c_{voll}+\delta-c_r\right)\left(a_{3t}-\overline{a}_{3t}\right)\Pr_{3t}\right] q_r.
\end{aligned}
$$

For *corr*, we have $a_{ct} = \overline{a}_{ct}$ so that $E\left[\pi_r\right] = 0$, as found in the one period setup. For *ind*, it is more complicated.

Define $\frac{\sum_t a_{2t}\Pr_{2t}}{\sum_t \Pr_{2t}} =: \alpha_2$, $\frac{\sum_t a_{3t}\Pr_{3t}}{\sum_t \Pr_{3t}} := \alpha_3$, and

$$
\Gamma := \begin{cases} 0 & \text{for } corr,\\ \frac{a-\alpha_2}{\alpha_2}\frac{b_r}{a} + \frac{\alpha_2-\alpha_3}{\alpha_2}\left(c_{voll}+\delta-c_r\right)\sum_t \Pr_{3t} & \text{for } ind. \end{cases}
$$

From the first-order condition of renewables firms, we obtain optimal probabilities of gas power dispatched (Ω_{2t}) for the respective cases,

$$\sum_t \Pr{}_{2t} = \begin{cases} \frac{b_r-(c_{voll}+\delta-c_r)\alpha_3\sum_t \Pr_{3t}}{(c_g-c_r)\alpha_2} & \text{for } corr, \\ \frac{b_r-(c_{voll}+\delta-c_r)a\sum_t \Pr_{3t}}{(c_g-c_r)a} & \text{for } ind. \end{cases}$$

Using this and α_2, α_3 in renewables firm's profit function yields

$$\begin{aligned} E[\pi_r] &= \left[\sum_t (c_g-c_r)(a_{2t}-a)\Pr{}_{2t} + \sum_t (c_{voll}+\delta-c_r)(a_{3t}-a)\Pr{}_{3t}\right] q_r \\ &= -\alpha_2 q_r \Gamma. \end{aligned} \tag{65}$$

G Proof of Lemma 4

Welfare from (1), depending on the possible intervals of events, changes to

$$W = -\sum_j b_j Q_j + \begin{cases} \sum_m U_m(D_m) - \sum_j c_j Y_j & \text{for } \tilde{Q}_r \in \Omega_1, \\ \sum_m U_m(D_m) - \sum_j c_j Y_j & \text{for } \tilde{Q}_r \in \Omega_2, \\ U_{high}(D_{high}) + U_{mid}(D_{mid}) + U_{low}\left(\sum_j Y_j - D_{high} - D_{mid}\right) & \\ -\sum_j c_j Y_j - (c_{L,low}+\delta)\left(\sum_m D_m - Y_j\right) & \text{for } \tilde{Q}_r \in \Omega_{3,low}, \\ U_{high}(D_{high}) + U_{mid}\left(\sum_j Y_j - D_{high}\right) - \sum_j c_j Y_j & \\ -(c_{L,low}+\delta) D_{low} - (c_{L,mid}+\delta)\left(D_{high}+D_{mid}-Y_j\right) & \text{for } \tilde{Q}_r \in \Omega_{3,mid}, \\ U_{high}\left(\sum_j Y_j\right) - \sum_j c_j Y_j - (c_{L,low}+\delta) D_{low} & \\ -(c_{L,mid}+\delta) D_{mid} - (c_{L,high}+\delta)\left(D_{high}-Y_j\right) & \text{for } \tilde{Q}_r \in \Omega_{3,high}. \end{cases} \tag{66}$$

Recapitulate that $Y_s = Q_s$ for all events, $Y_g = Q_g$ in lost load events, $Y_g = \sum_m D_m - Q_s - \tilde{Q}_r$ in gas power dispatched events, $Y_r = \sum_m D_m - Q_s$ when renewables are curtailed and $Y_r = \tilde{Q}_r$ else. Efficient demand as specified in Lemma 4 follows from differentiating (66) with respect to D_m. Observe that we can interchange differentiation and expectation again.

H Proof of Proposition 13

Depending on the realized state of the world, firms profits are

$$\pi_j = -b_j q_j + \begin{cases} \sum_m p_m \Delta_{j,m} + \rho_1 y_{j,eim} - c_j y_j & \Omega_1, \\ \sum_m p_m \Delta_{j,m} + \rho_2 y_{j,eim} - c_j y_j & \Omega_2, \\ p_{fwm}\Delta_{j,fwm} + p_{dam}\Delta_{j,dam} + p_{idm}\left(y_j - \Delta_{j,fwm} - \Delta_{j,dam} - y_{j,eim}\right) & \\ +\rho_{3,idm} y_{j,eim} - c_j y_j - \left(\tau_{P,idm} + \delta\right)\left(\sum_m \Delta_{j,m} + y_{j,eim} - y_j\right) & \Omega_{3,low}, \\ p_{fwm}\Delta_{j,fwm} + p_{dam}\left(y_j - \Delta_{j,fwm} - y_{j,eim}\right) + \rho_{3,dam} y_{j,eim} & \\ -c_j y_j - \left(\tau_{P,dam} + \delta\right)\left(\Delta_{j,fwm} + \Delta_{j,dam} + y_{j,eim} - y_j\right) - \left(\tau_{P,idm} + \delta\right)\Delta_{j,idm} & \Omega_{3,mid}, \\ p_{fwm}\left(y_j - y_{j,eim}\right) + \rho_{3,fwm} y_{j,eim} - c_j y_j - \left(\tau_{P,fwm} + \delta\right)\left(\Delta_{j,fwm} + y_{eim} - y_j\right) & \\ -\left(\tau_{P,dam} + \delta\right)\Delta_{j,dam} - \left(\tau_{P,idm} + \delta\right)\Delta_{j,idm} & \Omega_{3,high}. \end{cases}$$

Start with renewables firms and Ω_1: Only renewables firms produce in Ω_1 but total renewables output is curtailed due to excess capacity. Consequently, there are no restrictions on $y_{r,eim}$ and final production follows from $y_r = \sum_m \Delta_{r,m} + y_{r,eim}$. In Ω_2, all renewables firms produce with full capacity, $y_r = \tilde{q}_r$, and offer the difference between delivery pledges and production as imbalance energy, i.e., $y_{r,eim} = \tilde{q}_r - \sum_m \Delta_{r,m}$. In all remaining events, $y_r = \tilde{q}_r$ and $y_{r,eim}$ is unrestricted. We can now derive first-order conditions of renewables firms w.r.t. $\Delta_{r,m}$, i.e.,

$$\begin{aligned} \frac{\partial E\left[\pi|\Omega_{idm}\right]}{\partial \Delta_{idm}} &= p_{idm}\Pr{}_{12} - c_r \Pr{}_1 - \rho_2 \Pr{}_2 \\ &\quad - \left(\tau_{P,idm} + \delta\right)\Pr{}_{3,idm} - \left(\tau_{P,idm} + \delta\right)\Pr{}_{3,dam} - \left(\tau_{P,idm} + \delta\right)\Pr{}_{3,fwm}, \\ \frac{\partial E\left[\pi|\Omega_{idm}\right]}{\partial \Delta_{dam}} &= p_{dam}\left(\Pr{}_{12} + \Pr{}_{3,idm}\right) - c_r \Pr{}_1 - \rho_2 \Pr{}_2 \\ &\quad - \left(\tau_{P,idm} + p_{idm} + \delta\right)\Pr{}_{3,idm} - \left(\tau_{P,dam} + \delta\right)\left(\Pr{}_{3,dam} + \Pr{}_{3,fwm}\right), \\ \frac{\partial E\left[\pi|\Omega_{idm}\right]}{\partial \Delta_{fwm}} &= p_{fwm}\left(\Pr{}_{12} + \Pr{}_{3,idm} + \Pr{}_{3,dam}\right) - c_r \Pr{}_1 - \rho_2 \Pr{}_2 - p_{idm}\Pr{}_{3,idm} - p_{dam}\Pr{}_{3,dam} \\ &\quad - \left(\tau_{P,idm} + \delta\right)\Pr{}_{3,idm} - \left(\tau_{P,dam} + \delta\right)\Pr{}_{3,dam} - \left(\tau_{P,fwm} + \delta\right)\Pr{}_{3,fwm}. \end{aligned}$$

Binding first-order conditions yield

$$\begin{aligned} p_{idm} &= c_r \Pr{}_1 + c_g \Pr{}_2 + \left(\tau_{P,idm} + p_{idm} + \delta\right)\Pr{}_3, \\ p_{dam} &= c_r \Pr{}_1 + c_g \Pr{}_2 + \left(\tau_{P,idm} + p_{idm} + \delta\right)\Pr{}_{3,idm} + \left(\tau_{P,dam} + p_{dam} + \delta\right)\left(\Pr{}_{3,dam} + \Pr{}_{3,fwm}\right), \\ p_{fwm} &= c_r \Pr{}_1 + c_g \Pr{}_2 + \left(\tau_{P,idm} + p_{idm} + \delta\right)\Pr{}_{3,idm} \\ &\quad + \left(\tau_{P,dam} + p_{dam} + \delta\right)\Pr{}_{3,dam} + \left(\tau_{P,fwm} + p_{fwm} + \delta\right)\Pr{}_{3,fwm}. \end{aligned}$$

Using $\tau_{P,fwm} = c_{L,high,}$ in forward markets, $\tau_{P,dam} = c_{L,mid}$ in day-ahead markets, and $\tau_{P,idm} = c_{L,low}$ yields the wanted results.