On the efficient market diffusion of intermittent renewable energies*

Carsten Helm[†] *and Mathias Mier*[‡]

March 27th, 2018

Abstract

Capacity costs of renewable energies have been decreasing dramatically and are expected to fall further, making them more competitive with fossils. However, building on an analytically tractable peak-load pricing model we show that intermittency of renewable energies substantially impairs their market diffusion. In particular, once renewables have become competitive by attaining the same levelized cost of electricity (LCOE) as fossils, the marginal increase in efficient capacities due to a further cost reduction varies substantially. Initially it is small, then it rises convexly, but falls substantially once renewable capacities are large enough to satisfy the whole electricity demand at times of high availability. If external costs of fossils are internalized by a Pigouvian tax, then perfect competition leads to efficient investments in renewable and fossil capacities; even though we assume that only a subgroup of consumers can adapt their demand to price fluctuations that are caused by the intermittency of renewables. Maximum electricity prices rise with the share of renewables. If regulators impose a price cap, this initially raises investments in renewables, but the effect may reverse if the share of renewables is large.

Keywords: renewable energies, peak-load pricing, intermittent energy sources, technology diffusion, price caps, energy transition

JEL Classification: Q21, Q41, Q42, O33, D24

1 Introduction

When the G7 leading industrial nations agreed at their 2015 meeting to phase out the use of fossil fuels by the end of the century, the Economist wrote that: "In just a few years, the aim of a carbon-free energy system has gone from the realms of green fantasy to become official policy in

*This is a substantially revised version of C. Helm und M. Mier, "Efficient diffusion of renewable energies: A roller-coaster ride," University of Oldenburg, Department of Economics, Working Papers V-389-16, 2016. A previous version was circulated under the title "Efficient diffusion of intermittent renewable energies: A roller-coaster ride". We gratefully acknowledge helpful comments on this and the current version from participants at the Annual Conference of the Association for Public Economic Theory 2015, FSR Climate Annual Conference 2016, Conference of the European Association of Environmental and Resource Economists 2016, Conference of the Verein für Socialpolitik 2016, Workshop on Global Environmental Challenges at the Centre for European Economic Research (ZEW), as well as at research seminars at Darmstadt University of Technology, VATT Institute for Economic Research in Helsinki, Mercator Research Institute on Global Commons and Climate Change (MCC) in Berlin, and Hamburg University.

†Corresponding author. University of Oldenburg, Department of Business Administration, Economics and Law, 26111 Oldenburg, Germany, carsten.helm@uni-oldenburg.de, ph +49 441 798-4113.

‡University of Oldenburg, Department of Business Administration, Economics and Law, 26111 Oldenburg, Germany, mathias.mier@uni-oldenburg.de ph +49 441 798-4108.

the world's richest countries."[1] The adoption of the 2015 Paris agreement, in which 195 countries committed to the need for deep reductions in greenhouse gas emissions, extended this momentum to a global level. At the same time, the costs of renewable energies are decreasing rapidly. Within the period 2009-2014, the costs of solar PV modules fell by three-quarters, and those of wind turbines by almost a third (IRENA 2015). Further reductions between a quarter and around two-thirds are expected until 2025 (IRENA, 2016). This suggests that the economic viability of a transition from fossil to renewable energies is also improving. However, especially wind and solar power—the fastest-growing renewable energies—are peculiar products. Not only are they non-storable (at reasonable costs); their supply is also intermittent as it depends heavily on the fluctuations of wind speeds and solar radiation. We show that this has substantial implications for the efficient market diffusion of renewable energies.

We build on the peak-load pricing model, to analyze capacity investments and production decisions for an economy in which electricity can be produced from dispatchable fossil and intermittent renewable energies. Thus, in contrast to the standard peak-load pricing literature that considers variable demand (see Crew, Fernando, and Kleindorfer (1995) for a survey), the magnitude of intermittency follows endogenously from investments in renewable capacities. As pointed out by Ambec and Crampes (2012), renewables are always dispatched first due to their lower operating costs so that their intermittent supply causes an intermittent residual demand for fossils. Nevertheless, intermittent demand has quite different effects on investment incentives than intermittent supply of renewable energies. In the standard peak-load pricing model, the technology with lower variable costs (in our case renewables) produces at full capacity when demand and, therefore, prices are high. By contrast, if the same prices are caused by an equivalent low supply of renewables, these now have less to sell whereas the competing fossil energies take over and benefit from the high price. Conversely, when supply of renewable energies is high, the electricity price is low. As the share of renewables in the energy system rises, there will be extended periods in which they can meet all of the demand for electricity. As a result, prices drop to the level of the short-term operating costs for renewables, which are essentially zero.

Intuitively, the potential of renewables to earn profits is reduced by this price pattern. However, we show that this is a reflection of their market value so that competitive markets still lead to efficient choices of renewable and fossil capacities, provided that environmental costs are internalized by an appropriate tax. This is the case even if dynamic pricing of electricity is restricted to a subset of consumers that is equipped with the required technologies such as smart metering system; while the other consumers lack such technologies and, therefore, receive long-term fixed price contracts (Proposition 1). This contrasts with the findings in Ambec and Crampes (2012, Proposition 4) that competitive markets do generally not implement efficient capacity levels if (all) consumers receive fixed price contracts.

We then examine the efficient market diffusion of renewables as their capacity costs fall over time. Our focus lies on the effects of intermittency. Hence the model is constructed such that if renewables were not intermittent, they would completely replace investments in fossil capacities once having obtained the same levelized cost of electricity (LCOE). Accordingly, we deliberately abstract from other obstacles of technology diffusion (see Geroski 2000), and for our analysis it is also not relevant why renewables get cheaper, e.g. due to learning by doing or economies of scale.

We model intermittency by assuming that the availability of installed renewable capacities is a continuous random variable with a known distribution. Prices depend on the realization of this random variable and on the share of renewables in the energy systems, which determines their impact on the overall energy market. More specifically, when the market share of renewables is low, prices always exceed the marginal production costs of fossil energies and, as a result, renewable and fossil capacities are fully used. As the market share of renewables rises, there will be extended periods in which prices equal the marginal costs of fossils and fall below this level up to the marginal

[1] The Economist, "The G7 and climate change", June 10, 2015, http://www.economist.com/news/international/21653964-why-g7-talking-about-decarbonisation-sort.

production costs of renewables. This has profound implications for the efficient level of renewable and fossil capacities (Propositions 2 and 3). When renewables have become sufficiently cheap to compete with fossils (same LCOE), their efficient market share in the energy system initially grows only slowly in further cost reductions. Once the installed capacities are large enough to displace fossil production if availability of renewables is high, their efficient capacity level rises convexly in falling capacity costs. However, this strong increase falls substantially and turns from convex to concave once they can meet the entire energy demand at times of high availability. Intuitively, the price drop that results from further renewable capacities would then be borne primarily by the renewables themselves, and only to a lesser extent by fossils.

Roughly speaking, the market diffusion of renewable capacities follows an S-shaped pattern as their costs fall over time due to technological progress. This resembles the standard result in the innovation literature that the usage of new technologies over time typically follows an S-curve.[2] However, the mechanisms that lead to this result are quite different. In the most popular endemic model, it is the lack of information available about the new technology that hampers its market penetration. In our article, it is the intermittency of renewables.

The diffusion pattern of renewables is closely related to their competitiveness. This is usually assessed by comparing the levelized cost of electricity (LCOE) of renewables to that of conventional fossil technologies (e.g. IRENA (2016) and IEA (2015)). In our static framework, the LCOE would be defined as the *constant* price of power that equates expected revenues and costs. However, several authors have criticized the LCOE metric as flawed. Specifically, they point out that the market value of electricity varies widely over time, and that intermittent renewables cannot be dispatched when they would be most valuable (e.g. Joskow (2011) and Borenstein (2012)).[3] Instead of using the LCOE, Joskow (2011) suggests evaluating all technologies based on the expected market value of the electricity supplied, their total life-cycle costs, and their expected profitability. However, in our framework of competitive markets, this alternative is not very useful because in equilibrium, capacities are chosen such that all technologies are equally competitive. Although we agree that the LCOE is a flawed metric, we will argue that it can be quite useful in understanding the relative competitiveness of intermittent and dispatchable technologies, provided that it is interpreted appropriately. Moreover, by comparing the LCOE and market value of fossils we show that fossils get paid a markup through the market for offering back-up capacity that are needed to guarantee supply to consumers with fixed-price contracts.

At present, dynamic pricing of electricity is often restricted to larger commercial customers (Borenstein and Holland 2005, Joskow and Wolfram 2012). However, recent technological advances have dramatically lowered the costs of smart metering technologies, and many regions have set ambitious targets for their deployment.[4] This suggests that dynamic electricity pricing is likely to become more relevant for smaller commercial and residential customers too. Moreover, several studies have found evidence that households do actually respond to higher electricity prices by lowering usage (Jessoe and Rapson (2014), Faruqui and Sergici (2010)). The combination of marginal cost pricing and a higher share of intermittent renewables will lead to stronger price fluctuations and, in particular, to high maximum prices if availability is low. Policy makers may consider this politically unacceptable and impose a price cap in response. We find that this initially increases the competitiveness of renewables and supports their market diffusion. The reason is that fossils sell most of their output when prices are high; hence they are affected more severely by a price cap than renewables. However, this pattern is turned around once renewables have captured a large market share and fossils are mainly used as a back-up to guarantee that part of demand

[2] See Griliches (1957) for the seminal contribution and Geroski (2000) for a survey.

[3] Hirth (2013) highlights that the market value of intermittent renewable technologies is also affected by issues such as location (grid-related costs) and uncertainty (balancing costs), but these are not relevant in our simple analytical model.

[4] For example, the EU Third Energy Package requires Member States to ensure implementation of intelligent metering systems with a deployment target of at least 80 percent by 2020, conditional on a positive economic assessment of the long-term costs and benefits.

that cannot adapt to short-term price signals (Proposition 4).

The standard literature on the adoption of CO_2 abatement technologies like renewables energies has abstracted from the specific aspects of intermittency that are at the core of our contribution (e.g. Requate and Unold (2003), Fischer, Preonas, and Newell (2017)). Nevertheless, there exists by now a substantial literature that analyzes investment incentives with intermittent renewables. Most studies are either country-specific numerical simulations (e.g. Green and Vasilakos 2010) or empirical studies (e.g. Liski and Vehviläinen 2016). By contrast, the theoretical literature is still small—Ambec and Crampes (2012, p. 321) even wrote that "the economics of intermittent sources of electricity production are still in their infancy". Our article is the first that analytically determines the efficient market diffusion of renewables, with a focus on the effects of intermittency on diffusion stages with high shares of renewables, and allowing for dynamic short-term prices as well as long-term fixed price contracts. Ambec and Crampes (2012) also analyze the efficient mix of reliable and intermittent technologies as well as its decentralization by competitive markets. However, in their article, the availability of renewables is restricted to be either 0 or 1. Therefore, it is never efficient to build up capacities of renewables beyond the level at which they are used in state 1 of high availability, since these capacities would not be available in the other state, 0. Moreover, renewables and fossil capacities are either fully used or not at all. By contrast, in our article, the availability of renewables can take any positive value between 0 and 1. Therefore, it is often efficient to build up renewable capacities that lie idle for high values of availability but are used for lower values, and vice versa for fossil capacities. Such periods of excess capacity are crucial for the pattern of market penetration with renewables, which is the focus of our contribution. They also explain prices that equal the (very low) marginal cost of renewables that obtain in our model but not in Ambec and Crampes (2012).

Most other theoretical contributions have focused on public policies to promote renewables. Ambec and Crampes (2017) build on their earlier article with an analytical comparison of carbon taxes, feed-in tariffs, and renewable portfolios (see also Garcia, Alzate, and Barrera (2012)). As in our contribution, they distinguish between consumers with fixed-prices and state-contingent prices. Fell and Linn (2013) as well as Abrell, Rausch, and Streitberger (2018) both contain a simulation and a simpler analytical model with two renewable technologies (wind and solar) whose availability is either 0 or 1, but in different periods. Fell and Linn (2013) examine how this heterogeneity in the intermittency affects the cost effectiveness of various policies. In contrast to our model, demand is perfectly inelastic. Abrell et al. (2018) determine optimal renewable energy support policies. Andor and Voss (2016) are more similar to our model in that they allow the availability of renewables to take any value in the interval $[0, 1]$. However, their model does not include a second, fossil technology, and their focus lies on efficient subsidy schemes. In many respects, the model in Green and Léautier (2017) is the most general one of the theoretical contributions, incorporating several fossil and renewable technologies as well as intermittent supply and demand. Their focus lies on cost reductions of renewables that arise from higher installed capacities due to learning by doing and economies of scale. Formally, this is implemented by assuming a concave capacity cost function. Specifically, they analyze the costs—respectively the required subsidy (implemented by a feed-in tariff)—of attaining exogenously given capacity targets. Our analysis takes the reverse perspective. We determine the efficient capacity level for exogenously given constant marginal costs. Moreover, their analysis contains only reactive consumers and, in particular, focuses on small to medium shares of renewable energies (calibrated numerical simulations for market shares of electricity from renewables below 50 per cent), whereas the effects of excess capacities are crucial for our model. Their main result is that the need to subsidize renewables may never stop despite falling capacity costs because their market value drops even faster due to a discontinuity in the falling marginal value of renewables. This somewhat mirrors our results, where the lower market value of renewables reduces the effect of a further cost reduction on their efficient capacity level.

Other related theoretical contributions are Twomey and Neuhoff (2010) as well as Rouillon (2015). The former takes the capacity of the intermittent technology as given, and the latter the level of the reliable technology. By contrast, the feed-back effects from capacity investments in one technology on the incentives to invest in the other one feature prominently in our article. In addition, the focus of the two articles lies on optimal intermittent generation decisions in markets with market power. Chao (2011) considers uncertain supply of conventional and renewable technologies, where the main difference between the two is that demand is (negatively) correlated with supply from renewables, but uncorrelated with supply from conventional energies. More generally, the focus of the literature on supply (and demand) uncertainty lies on outage costs and rationing rules (see Kleindorfer and Fernando 1993). We ignore these complications for parsimony, but also because there has been tremendous progress in the reliability of forecast models (Iversen, Morales, Møller, and Madsen, 2016). Moreover, Gowrisankaran, Reynolds, and Samano (2016) who empirically estimate the effects of intermittency for southeastern Arizona, find that social costs of unforecastable intermittency are small in comparison to those of intermittency overall.

Finally, our article is related to the literature on price caps. Joskow and Tirole (2007) mention regulatory opportunism as one motivation for imposing caps on prices so as to keep them low. Stoft (2003) argues that price caps are useful to reduce market power and price volatility. Fabra, Von der Fehr, and De Frutos (2011) focus on market design and investment incentives. For a single-technology duopoly model of energy production they find that a price cap leads to underinvestment. In our analysis with competitive firms and two technologies, a price cap leads to investments below the efficient level for one technology, and to investment above the efficient level for the other one.

The remainder of the article is organized as follows. In the next section, we introduce the model. Sections 3 and 4 derive efficient production decisions and capacity choices for a renewable and a fossil technology. In Section 5, we show that the efficient solution can be implemented by competitive markets. Section 6 examines how falling capacity costs affect the market diffusion of renewables, and Section 7 considers the effects of a price cap. Section 8 concludes and an appendix contains all proofs.

2 The model

Consider a market in which electricity can be generated from two technologies, $j = r, f$. Technology f represents a dispatchable fossil technology—like conventional power plants that burn coal or gas. Dispatchability means that electricity production can be freely varied at every point in time up to the limit of their installed capacity (see Joskow 2011). Thus, for parsimony we ignore ramp-up times. Technology r is a renewable technology with intermittent supply—like wind turbines, solar PV and solar thermal plants.

The literature uses different assumptions about capacity costs. Green and Léautier (2017) argue that learning and economies of scale lead to concave cost functions, while Abrell et al. (2018) emphasize that investments first take place in the most productive sites, resulting in convex costs. We stay in between these two approaches and assume constant costs, $\beta_j > 0$, of providing one unit of capacity, $Q_j \geq 0$. Intermittency is represented by an availability factor, $\alpha \in [a, 1]$, where $0 < a < 1$. Thus, a higher a can be interpreted as a higher reliability of the renewable technology.[5] $F(\alpha)$ is the cumulative distribution function of α and $f(\alpha)$ its density. For substantial parts of the article, we assume a uniform distribution, $f(\alpha) = 1/(1-a)$, as this keeps the analysis tractable. In conclusion, the available capacity is αQ_r for renewables and Q_f for fossils. Finally, we denote by $b_r \geq 0$ and $c_f \geq 0$ the constant costs of producing one unit of output, $q_j \geq 0$, using the renewable and fossil technology, respectively. Only fossils lead to an environmental unit cost, $\delta \geq 0$, so that total costs of the fossil technology per unit of output are $b_f = c_f + \delta$. We assume that renewables

[5] In Germany, the minimum availability of installed wind and solar capacities in 2015 was 0.412 percent (own calculations based on hourly extrapolated data from the four German transmission system operators, downloaded from www.netztransparenz.de on 22 November 2016). However, such situations of extremely low availability are very rare and could be bridged, e.g. by better storage capabilities. Also more stable offshore wind power should lead to a higher a.

have lower variable costs than fossils, $0 \le b_r < b_f$.

In line with the literature on peak-load pricing, we consider one period that corresponds to the lifetime of installed capacities (assumed to be the same for all technologies). Thus, we abstract from issues of discounting and the complex dynamics that arise when new plants are built in addition to existing ones. We distinguish between reactive and non-reactive consumers that differ in their ability to adjust their electricity consumption in response to state contingent prices. We assume that this results from exogenous technological conditions—such as a different equipment with smart metering technologies—that equally constrain the first-best and the competitive solution. Let x_v denote electricity demand of reactive consumers, and $p_v(x_v)$ their inverse demand function. Hence their "gross surplus" in a particular state α is $\int_0^{x_v(\alpha)} p_v(\tilde{x}_v)\, d\tilde{x}_v$, where demand of reactive consumers, $x_v(\alpha)$, depends on the availability of renewable energies in state α. Similarly, x_n denotes electricity demand of non-reactive consumers, and $p_n(x_n)$ their inverse demand function. Hence their "gross surplus", which is independent of the particular state α, is $\int_0^{x_n} p_n(\tilde{x}_n)\, d\tilde{x}_n$.

Accounting for both groups of consumers, subtracting variable and fixed costs of the two technologies as well as environmental costs of fossils, and integrating over the different states of α yields expected welfare, W, as the sum of expected consumer and producer surplus minus environmental costs:

$$W = \int_a^1 \left[\int_0^{x_v(\alpha)} p_v(\tilde{x}_v)\, d\tilde{x}_v - \sum_j b_j q_j(\alpha) \right] dF(\alpha) + \int_0^{x_n} p_n(\tilde{x}_n)\, d\tilde{x}_n - \sum_j \beta_j Q_j, \quad (1)$$

where $b_f q_f = (c_f + \delta)\, q_f$ is the sum of production and environmental costs of fossils.

For parsimony, we assume linear demand functions $p_v(x_v) = \frac{\nu A - x_v}{\nu\gamma}$ for reactive consumers and $p_n(x_n) = \frac{(1-\nu)A - x_n}{(1-\nu)\gamma}$ for non-reactive consumers, so that $x_v = \nu(A - \gamma p_v)$ and $x_n = (1-\nu)(A - \gamma p_n)$. Thus, if both consumers faced the same price $p := p_n = p_v$, aggregate demand would be $x_v + x_n = A - \gamma p$, of which reactive and non-reactive consumers would receive the shares $\nu \in (0,1)$ and $1-\nu$, respectively. In this sense, ν and $1-\nu$ correspond to the shares of the two groups in the population.[6] Finally, in order to ensure that it is always efficient to install a positive capacity level, we assume that the maximum willingness to pay (WTP) exceeds the total costs per unit of fossils, i.e. $WTP_{max} = \frac{A}{\gamma} > b_f + \beta_f$.

The timing is as follows: In stage 1, a regulator chooses optimal capacities of renewables and fossils as well as electricity consumption of non-reactive consumers, based on the known distribution of the availability of renewables, α. Capacities must be chosen such that demand of non-reactive consumers can be matched for all realizations of α, which requires $aQ_r + Q_f \ge x_n$. Thus, we do not allow blackouts or load rationing. Note that the availability of renewables is intermittent, but there is no uncertainty in our model. Therefore, any blackouts would be planned, and at least for developed countries the political costs appear to be very high, given the great appreciation of a reliable electricity supply. In stage 2, the regulator chooses optimal production of renewables and fossils for a specific realization of α. By backwards induction, we first examine the second stage.

3 Production decisions and consumption of reactive consumers

In stage 2, capacities, $\mathbf{Q} := (Q_r, Q_f)$, are already installed and demand from non-reactive consumers, x_n, is given. The regulator chooses production, $\mathbf{q} := (q_r, q_f)$, and consumption of reactive

[6] The analysis can be extended straightforwardly to the (unrealistic) extreme cases of $\nu = 0$ and $\nu = 1$, by dropping all terms related to p_n, x_n or p_v, x_v, respectively. However, the notational costs in terms of additional case distinctions would be substantial so that we skipped this. A previous version of this paper considered the case of only reactive consumers ($\nu = 1$); it is available at https://www.uni-oldenburg.de/fileadmin/user_upload/wire/fachgebiete/vwl/V-389-16.pdf.

consumers, x_v, for a given availability of renewables, α. He does so to maximize the difference between gross consumer surplus, and variable production and environmental costs, subject to the constraints that supply equals demand and that supply from technology j cannot exceed the available capacity of this technology.

The gross surplus from non-reactive consumers, $\int_0^{x_n} p_n(\tilde{x}_n)d\tilde{x}_n$, is given at stage 2 and, therefore, not included in the decision problem. Denoting the value function of the stage 2 problem by $w(\mathbf{Q}, \alpha, x_n)$, we have

$$w(\mathbf{Q}, \alpha, x_n) := \max_{\mathbf{q}, x_v} \quad \int_0^{x_v} \tfrac{\nu A - \tilde{x}_v}{\nu\gamma} d\tilde{x}_v - \textstyle\sum_j b_j q_j \quad \text{such that} \tag{2}$$

$$\begin{aligned} \sum_j q_j - x_v - x_n &= 0, &(3)\\ \alpha Q_r - q_r &\geq 0, &(4)\\ Q_f - q_f &\geq 0. &(5) \end{aligned}$$

In addition, the program must satisfy the non-negativity constraints, $q_j \geq 0$ for $j = r, f$ and $x_v \geq 0$. However, they can be ignored because the solution of the problem without these constraints will never involve negative quantities due to our assumption that the maximum WTP exceeds variable costs. The Kuhn-Tucker Lagrangian is

$$\begin{aligned} \mathcal{L}(\mathbf{q}, x_v) &= \int_0^{x_v} \frac{\nu A - \tilde{x}_v}{\nu\gamma} d\tilde{x}_v - \sum_j b_j q_j + \lambda \left(\sum_j q_j - x_v - x_n \right) \\ &\quad + \mu_r (\alpha Q_r - q_r) + \mu_f (Q_f - q_f), \end{aligned} \tag{6}$$

where λ, μ_r, and μ_f are the multipliers for the supply-equals-demand and capacity constraints, respectively. The first-order conditions are (superscript * denotes efficient levels)

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_v} &= \tfrac{\nu A - x_v}{\nu\gamma} - \lambda \leq 0 \quad [= 0 \text{ if } x_v^* > 0], &(7)\\ \frac{\partial \mathcal{L}}{\partial q_r} &= -b_r + \lambda - \mu_r \leq 0 \quad [= 0 \text{ if } q_r^* > 0], &(8)\\ \frac{\partial \mathcal{L}}{\partial q_f} &= -b_f + \lambda - \mu_f \leq 0 \quad \left[= 0 \text{ if } q_f^* > 0\right]. &(9) \end{aligned}$$

Together with the complementary slackness conditions,

$$\begin{aligned} \mu_r \geq 0, \quad \mu_r [\alpha Q_r - q_r] &= 0, &(10)\\ \mu_f \geq 0, \quad \mu_f [Q_f - q_f] &= 0, &(11) \end{aligned}$$

this determines the endogenous variables, x_v, q_r, and q_f as functions of installed capacities, $\mathbf{Q}$, the availability of renewables, α, and demand of non-reactive consumers, x_n.

Several outcomes can be distinguished. First, stage 1 may have led to only fossil capacities (called *diffusion stage F*), only renewable capacities (*diffusion stage R*), or capacities of both types (*diffusion stage FR*). Second, we will show that in the two diffusion stages with renewables, in equilibrium four different cases may obtain that depend on the realization of the availability factor, α.

We now analyze the different outcomes, starting in a situation with low available capacities of renewables and then turning to those with higher levels. Thus, we first consider diffusion stage *F,* in which only fossil capacities have been built. Given that neither supply nor demand are intermittent in this diffusion stage, the distinction between reactive and non-reactive is irrelevant so that we use p and x to denote the electricity price and aggregate demand. Obviously, excess capacities would never be used, so it cannot be efficient to install them. Therefore, $x^* = q_f^* = Q_f > 0$ and the price follows from the specification of inverse demand as $p^* = (A - Q_f)/\gamma$.

Next, consider diffusion stage *FR,* where fossil and renewable technologies have been installed. We focus on a graphical exposition and relegate the formal analysis to Appendix A. Figure 1 depicts (inverse) demand of reactive consumers and four supply functions that correspond to different levels of available renewable capacities, αQ_r. Renewables have lower production costs and are therefore always dispatched first. Accordingly, in all four cases, the supply curve starts with a horizontal segment at the level of variable costs of renewables, b_r. Once the available renewable capacity is fully used, the supply curve jumps to the level of variable costs of fossils, b_f. Once also the fossil capacity is fully used, the supply curve is vertical. Remember that the first x_n units of supply are needed to satisfy demand of non-reactive consumers. Therefore, only the residual supply is available for reactive consumers and depicted in the positive quadrant.

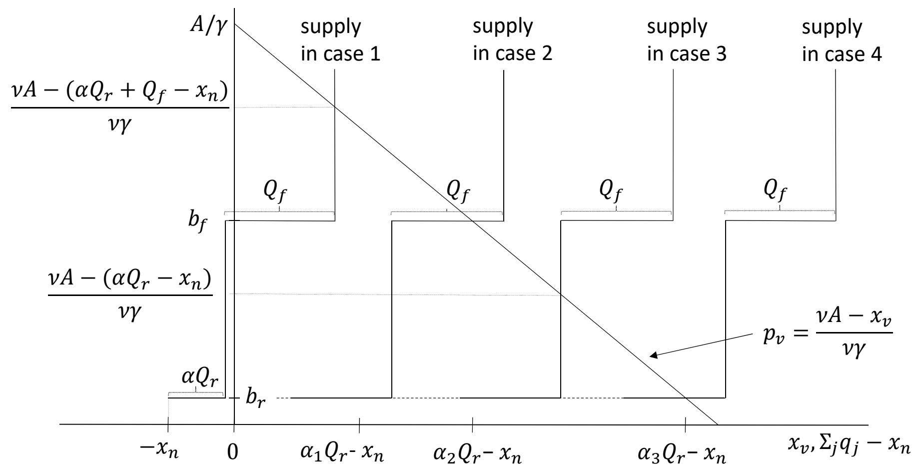


Fig. 1: Equilibrium on electricity market

Case 1 (we denote equilibrium values in the four different cases with subscript $i = 1, ..., 4$) refers to the situation where αQ_r is low so that the intersection with the (inverse) demand curve occurs in the last, vertical segment of the supply curve. Thus, both (available) capacities are fully used, i.e. $q_{r1}(\alpha) = \alpha Q_r$ and $q_{f1}(\alpha) = Q_f$, whereas the price follows immediately from the specification of inverse demand (see first line of Table 1). An increase in αQ_r shifts the supply curve to the right and reduces the equilibrium price until the upper horizontal segment of the supply curve starts to intersect with the demand curve, i.e. until $p_{v1}(\alpha) = (\nu A - \alpha Q_r - Q_f + x_n)/\nu\gamma = b_f$. Solving for α yields the first cut-off point, denoted $\alpha_1 := min\{(\nu A - \nu\gamma b_f - Q_f + x_n)/Q_r, 1\}$, such that case 1 obtains for all $\alpha \leq \alpha_1$. This definition of α_1 takes into account that the upper bound of the support of $f(\alpha)$ is 1.

As αQ_r rises further, the supply and demand curve continue to intersect at the variable cost of fossils, b_f (case 2). Thus, neither the equilibrium price nor demand change. However, production from renewables successively replaces production from fossils, leading to increasing excess capacities of the latter. The respective values follow straightforwardly from Figure 1 and $q_f = x_n + x_v - \alpha Q_r$. They are stated in the second line of Table 1. This case continues until available renewable capacities are sufficient to satisfy the entire demand of reactive and non-reactive consumers at $p_v = b_f$, i.e. until $\alpha Q_r = v(A - \gamma b_f) + x_n$. This yields the second cut-off point, $\alpha_2 := min\{(\nu(A - \gamma b_f) + x_n)/Q_r, 1\}$, and case 2 obtains for all $\alpha \in (\alpha_1, \alpha_2]$.

Tab. 1: Distinction of cases for $Q_r > 0$

i	availability	$p_{vi}(\alpha)$	$x_{vi}(\alpha)$	$q_{ri}(\alpha)$	$q_{fi}(\alpha)$
1	$a \leq \alpha \leq \alpha_1$	$\frac{\nu A-(\alpha Q_r+Q_f-x_n)}{\nu\gamma}$	$\alpha Q_r + Q_f - x_n$	αQ_r	Q_f
2	$\alpha_1 < \alpha \leq \alpha_2$	b_f	$\nu(A-\gamma b_f)$	αQ_r	$\nu(A-\gamma b_f) - \alpha Q_r + x_n$
3	$\alpha_2 < \alpha \leq \alpha_3$	$\frac{\nu A-(\alpha Q_r-x_n)}{\nu\gamma}$	$\alpha Q_r - x_n$	αQ_r	0
4	$\alpha_3 < \alpha \leq 1$	b_r	$\nu(A-\gamma b_r)$	$\nu(A-\gamma b_r)+x_n$	0

For further increases of αQ_r, demand intersects with the lower vertical segment of the supply curve (case 3). Now, the equilibrium price falls again in αQ_r, and the equilibrium values as given in the third line of Table 1 follow immediately from Figure 1. This case obtains until the available renewable capacity equals aggregate demand at the variable cost of renewables, b_r, which defines the third cut-off point $\alpha_3 := min\{(\nu(A - \gamma b_r) + x_n)/Q_r, 1\}$. For even higher values of αQ_r, there are excess capacities of renewables, which is case 4 in Table 1. Finally, in diffusion stage *R,* where only renewable capacities have been installed, the supply curve consists only of the lower horizontal and vertical segments. It follows immediately that case 3 occurs for all $\alpha \in [a, \alpha_3]$ and case 4 for all $\alpha \in (\alpha_3, 1]$, with quantities and prices as given in Table 1. Lemma 1 summarizes these results.

Lemma 1. *In diffusion stage* F *(i.e.* $Q_f > 0, Q_r = 0$*), equilibrium prices and quantities are* $x^* = q_f^* = Q_f$ *and* $p^* = (A - Q_f)/\gamma$*. In diffusion stage* FR *(i.e.* $Q_f, Q_r > 0$*), the solution depends on the availability of renewables,* α*, as summarized by the four cases in Table 1. In diffusion stage* R *(i.e.* $Q_f = 0, Q_r > 0$*), only cases 3 (for* $\alpha \in [a, \alpha_3]$*) and 4 in Table 1 obtain.*

4 Capacity choices and consumption of non-reactive consumers

We now turn to the regulator's choices of capacities and the fixed quantity supplied to non-reactive consumers, x_n. As above, we first consider diffusion stage F of fossils only, for which the distinction between reactive and non-reactive consumers is irrelevant. Efficiency requires that the equilibrium price equals fossil's long-run marginal costs, $b_f + \beta_f$. This yields demand, output, and optimal capacity $q_f^* = x^* = Q_f^* = A - \gamma(b_f + \beta_f)$. Moreover, suppose that the first marginal unit of renewables would be added to the system. Given its lower variable costs, it would always be employed. Assuming a uniform distribution of α, this leads to an expected output of $\frac{1+a}{2}$ and associated expected costs of $\frac{1+a}{2}b_r + \beta_r$. The costs of producing the same output by fossils are $\frac{1+a}{2}(b_f + \beta_f)$. Comparing costs, it is efficient to employ renewables if and only if $\beta_r \leq \frac{1+a}{2}(b_f - b_r + \beta_f) =: \overline{\beta_r}$. Intuitively, the critical capacity cost of renewables, $\overline{\beta_r}$, is higher the higher the capacity cost of fossils, the larger the difference in the variable costs of fossils and renewables, and the better the reliability of renewables, a.

Next, consider diffusion stages FR and R for which intermittency of renewables has to be taken into account. Remember that we have denoted the difference between the gross surplus of reactive consumers and the sum of variable production and environmental costs in a particular state α that obtains from the optimization problem in stage 2 by $w(\mathbf{Q}, \alpha, x_n)$. Taking into account that prices and quantities vary over the support $[a, 1]$ of $F(\alpha)$, including the gross surplus of non-reactive consumers and accounting for capacity costs, the welfare maximization problem in stage 1 is

$$\max_{\mathbf{Q},x_n} W(\mathbf{Q}) = \int_a^1 w(\mathbf{Q}, \alpha, x_n)\, dF(\alpha) + \int_0^{x_n} \frac{(1-\nu)A - \tilde{x}_n}{(1-\nu)\gamma} d\tilde{x}_n - \sum_j \beta_j Q_j \qquad (12)$$

$$\text{such that} \quad aQ_r + Q_f \geq x_n, \qquad (13)$$

where the constraint assures that supply is always sufficient to satisfy demand of non-reactive consumers. The Kuhn-Tucker Lagrangian is

$$\begin{aligned} \widetilde{\mathcal{L}}(\mathbf{q}, x_v) &= \int_a^1 w(\mathbf{Q}, \alpha, x_n)\, dF(\alpha) + \int_0^{x_n} \frac{(1-\nu)A - \tilde{x}_n}{(1-\nu)\gamma} d\tilde{x}_n - \sum_j \beta_j Q_j \\ &\quad + \zeta(aQ_r + Q_f - x_n) \end{aligned} \tag{14}$$

where ζ is the Lagrangian multiplier. According to the Leibniz rule (e.g. Sydsaeter, Hammond, Seierstad, and Strom 2005, p. 156),

$$\frac{\partial}{\partial Q_j}\int_a^1 w(\mathbf{Q}, \alpha, x_n)\, dF(\alpha) = \int_a^1 \frac{\partial w(\mathbf{Q}, \alpha, x_n)}{\partial Q_j} dF(\alpha) \quad \text{for } j = r, f \tag{15}$$

and

$$\frac{\partial}{\partial x_n}\int_a^1 w(\mathbf{Q}, \alpha, x_n)\, dF(\alpha) = \int_a^1 \frac{\partial w(\mathbf{Q}, \alpha, x_n)}{\partial x_n} dF(\alpha) \tag{16}$$

if $w(\mathbf{Q}, \alpha, x_n)$, $\partial w(\mathbf{Q}, \alpha, x_n)/\partial Q_j$, and $\partial w(\mathbf{Q}, \alpha, x_n)/\partial x_n$ are continuous. From Table 1 and equation (2), $w(\mathbf{Q}, \alpha, x_n)$ is obviously continuous *within* the four cases because prices and quantities are continuous within these cases. To see that it is also continuous at the boundaries of two neighboring cases, note that $w_i(\mathbf{Q}, \alpha_i, x_n) = w_{i+1}(\mathbf{Q}, \alpha_i, x_n), i = 1, 2, 3$ because prices and quantities are the same at these boundaries (see Table 1). Turning to the derivatives, from equation (2) we obtain

$$\frac{\partial w(\mathbf{Q}, \alpha, x_n)}{\partial Q_j} = \frac{\partial x_v}{\partial Q_j}\frac{\nu A - x_v}{\nu\gamma} - \sum_j b_j \frac{\partial q_j}{\partial Q_j}. \tag{17}$$

Using the values from Table 1, this is obviously continuous within the four cases. Moreover, $x_{v1}(\alpha_1) = \nu(A - \gamma b_f) = x_{v2}(\alpha_1)$, and $q_{r1}(\alpha_1) = \nu(A - \gamma b_f) - Q_f + x_n = q_{r2}(\alpha_1)$, and $q_{f1}(\alpha_1) = Q_f = q_{f2}(\alpha_1)$ so that at the border between cases 1 and 2, equation (17) becomes

$$\frac{\partial w_1(\mathbf{Q}, \alpha_1, x_n)}{\partial Q_f} = 0 + b_r - b_f = \frac{\partial w_2(\mathbf{Q}, \alpha_1, x_n)}{\partial Q_f}. \tag{18}$$

Using the same steps, it is straightforward to show that $\partial w(\mathbf{Q}, \alpha, x_n)/\partial Q_j$ and $\partial w(\mathbf{Q}, \alpha, x_n)/\partial x_n$ are also continuous at the respective borders of the other cases. Thus, we can apply (15) and (16) to determine the first-order conditions w.r.t. Q_f, Q_r, and x_n in diffusion stages *FR* and *R*. Splitting up the overall integral into the four different cases, differentiating $w(\mathbf{Q}, \alpha, x_n)$ as given in (2), and substituting the respective values from Table 1, we obtain the first-order partial derivatives of the Lagrangian.[7] Defining $g(Q_r, Q_f, x_n) := aQ_r + Q_f - x_n$ and using subscripts to denote derivatives, they can be stated as

$$W_{Q_f} + \zeta g_{Q_f} = \int_a^{\alpha_2} (p_v(\alpha) - b_f)\, dF(\alpha) - \beta_f + \zeta \le 0 \ \left[= 0 \text{ if } Q_f^* > 0\right], \tag{19}$$

[7] For case 1, differentiation w.r.t. Q_r and using $x_{v1} = \alpha Q_r + Q_f - x_n, q_{r1} = \alpha Q_r, q_{f1} = Q_f$ this yields

$$\int_a^{\alpha_1} \frac{\partial w}{\partial Q_r} dF(\alpha) = \int_a^{\alpha_1} \left(\frac{\partial x_v}{\partial Q_r}\frac{\nu A - x_v}{\nu\gamma} - b_r\frac{\partial q_r}{\partial Q_r} - b_f\frac{\partial q_f}{\partial Q_r}\right) dF(\alpha) = \int_a^{\alpha_1} (p_v(\alpha) - b_r)\, \alpha dF(\alpha).$$

The other derivatives are calculated along the same lines.

$$W_{Q_r} + \zeta g_{Q_r} = \int_a^1 \left(p_v\left(\alpha\right) - b_r\right) \alpha dF\left(\alpha\right) - \beta_r + a\zeta = 0 \ \left[\alpha_1 = \alpha_2 = a \text{ if } Q_f^* = 0\right], \quad (20)$$

$$W_{x_n} + \zeta g_{x_n} = p_n\left(x_n\right) - \int_a^1 p_v\left(\alpha\right) dF\left(\alpha\right) - \zeta \leq 0 \ \left[= 0 \text{ if } x_n^* > 0\right], \quad (21)$$

which defines the solution together with the complementary slackness condition

$$\zeta \geq 0, \ \zeta\left(aQ_r + Q_f - x_n\right) = 0. \quad (22)$$

The second-order sufficient condition requires that welfare $W\left(\mathbf{Q}\right)$ is concave and $\zeta\left(aQ_r + Q_f - x_n\right)$ is quasiconcave (see Sydsaeter et al. 2005, p. 134). The latter is trivially satisfied, and concavity of $W\left(\mathbf{Q}\right)$ is shown in the proof of Proposition 2.

The above first-order conditions take into account that $Q_f > 0$ in stage *FR* but $Q_f = 0$ in stage *R,* whereas $Q_r > 0$ in both stages. Intuitively, capacities Q_f and Q_r are chosen such that their respective marginal costs, β_j, are equal to their expected marginal value after accounting for production and environmental costs (the integral terms) and the shadow price, ζ, of the constraint that available capacities must be sufficient to meet demand of non-reactive consumers. The range of the integrals in condition (19) reflects that fossils only produce in cases 1 and 2.

Similarly, consumption of non-reactive consumers, x_n, is chosen such that the marginal gains of non-reactive consumers equal the marginal losses of reactive consumers—given by their respective (expected) prices—and the shadow price, ζ. The latter captures the effect that the constraint becomes more demanding as x_n increases. Intuitively, $x_n^* > 0$ because the maximum WTP of non-reactive consumers, $\frac{A}{\gamma}$, exceeds the price of reactive consumers, $\frac{\nu A - x_v}{\nu\gamma}$, for any $x_v > 0$. More formally, if $x_n = aQ_r + Q_f$, then $x_n^* > 0$ follows trivially from our assumption that $a > 0$. In the alternative case of $x_n < aQ_r + Q_f$, we have $\zeta = 0$ by the complementary slackness condition (22). At $x_n = 0$, the first-order derivative w.r.t. x_n becomes $\frac{(1-\nu)A - 0}{(1-\nu)\gamma} - \int_a^1 p_v\left(\alpha\right) dF\left(\alpha\right)$. However, this term is strictly positive since the maximum WTP, A/γ (the 1st term), obviously exceeds the expected price of reactive consumers (the 2nd term). This violates (21), and we conclude that the optimal solution has $x_n^* > 0$.

If the other two conditions (19) and (20) bind, we are in diffusion stage *FR.* If only condition (20) binds, we are in diffusion stage *R*, for which cases 1 and 2 can be dropped (hence $\alpha_1 = \alpha_2 = a$, see Lemma 1). Moreover, the transition from diffusion stage *FR* to *R* occurs when (19) binds at $Q_f = 0$. Hence, we can derive the capacity level and capacity costs of renewables where fossils are squeezed completely out of the market by solving the binding equation (19) at $Q_f = 0$ for Q_r, and then using this to solve (20) and the binding condition (21) at $Q_f = 0$ for β_r. We denote the resulting capacity levels and costs by $\underline{Q_r}$ and $\underline{\beta_r}$, but refrain from stating the specific values as they are of little relevance and very complex. The following lemma summarizes efficient capacity levels as a function of the capacity costs of renewables.

Lemma 2. *If capacity costs of renewables are $\beta_r \geq \overline{\beta_r}$, it is efficient to install only fossil capacities at the level $Q_f^* = A - \gamma\left(b_f + \beta_f\right)$. For $\beta_r \in \left(\underline{\beta_r}, \overline{\beta_r}\right)$, it is efficient to install fossil and renewable capacities at the levels that solve the system of binding equations (19) to (21). For $\beta_r \leq \underline{\beta_r}$, it is efficient to install only renewable capacities at the level that solves equations (20) and (21) with $\alpha_1 = \alpha_2 = a$.*

5 Market efficiency

In the preceding section, we determined welfare-maximizing levels of renewable and fossil capacities. We now ask whether this solution can also be achieved by decentralized markets if fossil firms face a Pigouvian tax $\tau = \delta$ that equals environmental cost per unit of output. We assume that there are a large number of competitive firms (indexed by superscript k) that produce with either the fossil or the renewable technology. In stage 1, a firm k that operates with technology $j = r, f$ installs capacity Q_j^k, yielding overall capacities $Q_j = \sum_k Q_j^k$. In stage 2, firms sign long-term fixed price contracts with non-reactive consumers about delivering x_{nj}^k at price p_n. This obliges each firm to set aside the necessary production capacities, i.e. $x_{nf}^k \leq Q_f^k$ and $x_{nr}^k \leq aQ_r^k$. In stage 3, the availability of renewables, α, realizes, electricity is produced and firms can use their remaining production capacities to sell $x_{vj}^k(\alpha)$ at state contingent prices $p_v(\alpha)$. We allow trading among firms. In particular, a firm k can meet its obligations under long-term contracts, x_{nj}^k, either by using its own capacities or by purchasing electricity at the state contingent price $p_v(\alpha)$. The latter is profitable for a fossil firm if $p_v(\alpha)$ is below its own production costs due to a high α.

By backwards induction, first consider stage 3, in which capacities and sales to non-reactive consumers, x_{nj}^k, are fixed and α has realized. Given competitive markets, the outcome is obviously efficient and leads to the results as summarized in Lemma 1 and Table 1. Turning to stage 2, we begin with the analysis of diffusion stage *FR*, in which fossil and renewable capacities have been installed. A fossil firm chooses x_{nf}^k so as to maximize its expected profits

$$\begin{aligned} E\left[\pi_f^k\right] &= \int_a^{\alpha_2} \left(p_v(\alpha) - b_f\right)\left(Q_f^k - x_{nf}^k\right) dF(\alpha) + \\ &\quad \left(p_n - \int_a^{\alpha_2} b_f dF(\alpha) - \int_{\alpha_2}^1 p_v(\alpha)\, dF(\alpha)\right) x_{nf}^k - \beta_f Q_f^k \text{ such that } Q_f^k \geq x_{nf}^k, \end{aligned} \quad (23)$$

where $b_f = c_f + \tau$ includes variable production costs, c_f, and the Pigouvian tax, $\tau = \delta$. Thus, the first integral term are expected net revenues from sales to reactive consumers, $x_{vf}^k = Q_f^k - x_{nf}^k$, which takes into account that fossil firms only produce for $\alpha \leq \alpha_2$ (see Table 1). The second term are expected net revenues from sales to non-reactive consumers at the price p_n. Here, firms anticipate that they will use their own production capacities if their variable production costs are weakly below the state contingent price $p_v(\alpha)$—which is the case for $\alpha \leq \alpha_2$—but buy electricity at the price $p_v(\alpha) < b_f$ for higher realizations of α. The first-order condition of the corresponding Kuhn-Tucker Lagrangian is

$$\frac{\partial E\left[\pi_f^k\right]}{\partial x_{nf}^k} = p_n - \int_a^1 p_v(\alpha)\, dF(\alpha) - \zeta_f \leq 0 \ \left[= 0 \text{ if } x_{nf}^{k*} > 0\right], \quad (24)$$

where $\zeta_f \geq 0$ is the Lagrangian multiplier. Similarly, renewable firms choose x_{nr}^k so as to maximize their expected profits

$$\begin{aligned} E\left[\pi_r^k\right] &= \int_a^1 \left(p_v(\alpha) - b_r\right)\left(\alpha Q_r^k - x_{nr}^k\right) dF(\alpha) \\ &\quad + \left(p_n - b_r\right) x_{nr}^k - \beta_r Q_r^k \text{ such that } aQ_r^k \geq x_{nr}^k. \end{aligned} \quad (25)$$

Here, the integral term are net revenues from sales to reactive consumers and fossil firms. Note that the state contingent price can never fall below the variable costs of renewable firms so that they always use their own capacities to produce x_{nr}^k. Denoting the Lagrangian multiplier by $\zeta_r \geq 0$, the first-order condition of the corresponding Kuhn-Tucker Lagrangian is

$$\frac{\partial E\left[\pi_r^k\right]}{\partial x_{nr}^k} = p_n - \int_a^1 p_v(\alpha)\, dF(\alpha) - \zeta_r \leq 0 \ \left[= 0 \text{ if } x_{nr}^{k*} > 0\right]. \quad (26)$$

We now show that $x_{nf}^{k*}, x_{nr}^{k*} > 0$ so that (24) and (26) bind. Given that we consider diffusion stage *FR* with $Q_r^k, Q_f^k > 0$, this is trivially the case if the constraints $x_{nf}^k \leq aQ_f^k$ and $x_{nr}^k \leq aQ_r^k$ bind. Next, suppose that at least one of the constraints does not bind; say $x_{nf}^k < aQ_f^k$ so that

$\zeta_f = 0$ by complementary slackness. If the first-order condition (24) does not bind, then also (26) does not bind (due to $\zeta_r \geq 0$) so that $x^k_{nf} = x^k_{nr} = 0$. However, in this case $p_n = \frac{(1-\nu)A-0}{(1-\nu)\gamma} = \frac{A}{\gamma}$ and, therefore, $p_n - \int_a^1 p_v(\alpha)\, dF(\alpha) > 0$, in violation of condition (24) at $\zeta_f = 0$. An equivalent argument applies if we start with $x^k_{nr} < aQ^k_r$. Hence we conclude that $x^{k*}_{nf}, x^{k*}_{nr} > 0$ so that (24) and (26) bind. This immediately implies $\zeta_f = \zeta_r$. Intuitively, as long as a firm's capacity constraint does not bind ($\zeta_f = 0$ or $\zeta_r = 0$), the (expected) prices of selling to non-reactive and reactive consumers must be the same.

In stage 1, firms anticipate that sales of x^k_{nj} follow from (24) and (26). Solving these expression for p_n and substitution into (23) and (25), expected profits become

$$E\left[\pi^k_f\right] = \int_a^{\alpha_2} \left(p_v(\alpha) - b_f\right) Q^k_f dF(\alpha) - \beta_f Q^k_f + \zeta_f x^k_{nf}, \tag{27}$$

$$E\left[\pi^k_r\right] = \int_a^{1} \left(p_v(\alpha) - b_r\right) \alpha Q^k_r dF(\alpha) - \beta_r Q^k_r + \zeta_r x^k_{nr}. \tag{28}$$

Noting that $\zeta_f = \zeta_r$, either both capacity constraints bind or none. In the first case, $\zeta_f = \zeta_r > 0$ as well as $x^k_{nf} = Q^k_f$ and $x^k_{nr} = aQ^k_r$, where the latter imply $x_n = \sum_j \sum_k x^k_{nj} = aQ_r + Q_f$. Upon comparing (21) with (24) and (26), we have $\zeta_f = \zeta_r = \zeta$ at efficient capacity levels. Substituting this together with $x^k_{nf} = Q^k_f$ and $x^k_{nr} = aQ^k_r$ into (27) and (28) shows that firms make zero profits at the efficient capacity levels that satisfy (19) and (20). It is straightforward to see that this is also the case if $\zeta_f = \zeta_r = \zeta = 0$.

Next, consider diffusion stage *F*. Given that neither supply nor demand are intermittent, fossil capacities are always fully dispatched, i.e. $q^k_f = Q^k_f$ and the electricity price, p, is constant. Thus, profits of a fossil firm are $\pi^k_f = (p - b_f - \beta_f)\, Q^k_f$. At efficient capacity levels, p equals fossil's long-run marginal costs, $b_f + \beta_f$, so that $\pi^k_f = 0$.

Finally, in diffusion stage *R,* only cases 3 and 4 are realized. Thus expected profits are as given by (28) after dropping cases 1 and 2 by setting $\alpha_1 = \alpha_2 = a$. Comparing this with (20), thereby distinguishing again the cases of a binding and non-binding capacity constraint, shows that profits of a renewable firm in diffusion stage *R* are zero at the efficient capacity levels. Hence there are again no incentives to enter or exit the market and we obtain the following result.

Proposition 1. *The efficient levels of fossil and renewable capacities can be implemented by competitive markets.*

6 The efficient market diffusion of renewables

Compared to fossils, renewables are still a new technology, for which falling capacity costs are anticipated (see IRENA 2016). In Lemma 2, we have already shown that renewables enter the market when capacity costs have fallen to the threshold value $\overline{\beta_r}$, and that they squeeze fossils completely out of the market when capacity costs are below $\underline{\beta_r}$. Now we take a closer look at the market diffusion in the intermediate stage *FR,* where fossil and renewable technologies coexist.

For this stage, we know from Section 4 that $\alpha_1 \leq \alpha_2 \leq \alpha_3$ and that each of these threshold values can be equal to 1. Thus, it may well be that only a subset of the cases in Table 1 obtains. However, some of them always do.

Lemma 3. *If it is optimal to install renewable and fossil capacities (diffusion stage* FR*), then capacity levels are chosen such that* $\alpha_1 > a$*. Hence there are always realizations of* α *for which both technologies are used at full capacity (case 1). If it is optimal to install only renewable capacities (diffusion stage* R*), then capacity levels are chosen such that* $\alpha_3 > a$*. Hence there are always realizations of* α *for which renewables are used at full capacity (case 3).*

The lemma reflects the intuitive idea that capacities are only installed if they are at least used for low realizations of α. Accordingly, in diffusion stage *FR,* only case 1 occurs if $\alpha_1 = 1$, which happens if $Q_r^* \leq \nu (A - \gamma b_f) - Q_f^* + x_n^*$ (see Table 1). Compared to the following situations, this is associated with the lowest level of renewables. Hence we refer to this diffusion stage as *very low renewables* (stage V). If $\alpha_1 < 1$ but $Q_r^* \leq \nu (A - \gamma b_f) + x_n^*$ so that $\alpha_2 = 1$, then only cases 1 and 2 obtain. We call this situation *low renewables* (stage L). Next, if $\alpha_2 < 1$ but $Q_r^* \leq \nu (A - \gamma b_r) + x_n^*$ so that $\alpha_3 = 1$, which allows for higher values of Q_r^* than the previous situation, then cases 1, 2, and 3 obtain. We call this situation *medium renewables* (stage M). Finally, if $\alpha_3 < 1$ all four cases obtain. We call this situation *high renewables* (stage H), which reflects that in case 4 there is excess capacity of renewables.

Intuitively, the efficient level of renewable capacities should be higher if they are cheaper. One would also expect that this goes along with a lower level of fossil capacities, but it turns out that this need not be the case. Consider first the case where the constraint $aQ_r + Q_f \geq x_n$ does not bind so that $\zeta = 0$ by complementary slackness. Thus, efficient capacity levels follow from the system of the three (binding) first-order conditions (19) to (21) for $\zeta = 0$. Applying the implicit function theorem to this equation system yields (for a uniform distribution of α)

$$\begin{pmatrix} \frac{\partial Q_r}{\partial \beta_r} \\ \frac{\partial Q_f}{\partial \beta_r} \\ \frac{\partial x_n}{\partial \beta_r} \end{pmatrix} = \begin{pmatrix} \frac{\alpha_1 - a}{\nu^2 \gamma^2 (1-a)^2 |A|} \left(\alpha_3 - \alpha_2 + \frac{\nu}{1-\nu} (1-a) \right) \\ \frac{\partial x_n}{\partial \beta_r} - \frac{1}{2} (\alpha_1 + a) \frac{\partial Q_r}{\partial \beta_r} \\ \frac{\alpha_1 - a}{\nu^2 \gamma^2 (1-a)^2 |A|} \frac{1}{2} \left(\alpha_3^2 - \alpha_2^2 \right) \end{pmatrix}, \tag{29}$$

where $|A| < 0$ is the determinant of the Hessian matrix (see Appendix C). It follows immediately that $\frac{\partial Q_r}{\partial \beta_r} < 0$ as expected. Next, $\frac{\partial x_n}{\partial \beta_r} = 0$ in stages V and L, for which $\alpha_3 = \alpha_2 = 1$. Intuitively, in these stages with low renewable capacities only cases 1 and 2 obtain in which fossil capacities are also used to serve demand from reactive consumers. Cheaper renewables replace fossil capacities at a level such that (expected) electricity prices and, therefore, demand of non-reactive consumers do not change. By contrast, in stages M and H for $\alpha \in (\alpha_2, \alpha_3]$ case 3 occurs, in which more renewable capacities reduce the expected electricity price for reactive consumers. Therefore, some of the additional electricity output is used to serve non-reactive consumers, i.e. $\frac{\partial x_n}{\partial \beta_r} < 0$, as reflected in equation (21) according to which the (expected) prices of reactive and non-reactive consumers must be equalized (at $\zeta = 0$).

Finally, consider the term $\frac{\partial Q_f}{\partial \beta_r}$ in (29). In stages V and L for which $\frac{\partial x_n}{\partial \beta_r} = 0$, fossil capacities are replaced by additional renewable capacities at the level of their expected availability in case 1, $\frac{1}{2} (\alpha_1 + a)$. Hence we obtain the expected sign $\frac{\partial Q_f}{\partial \beta_r} > 0$. By contrast, in stages M and H there is an opposite effect because demand of non-reactive consumers increases as renewables get cheaper. The effect is particularly strong if the share of non-reactive consumers is large (respectively, if ν is low). This additional demand must also be served if the availability of renewables is low and, thus, raises the price in case 1. This makes fossils more valuable, and if this effect dominates it may be efficient to increase fossil capacities in response to cheaper renewables. Formally, from (29) this happens, i.e. $\frac{\partial Q_f}{\partial \beta_r} < 0$, if and only if

$$\frac{1}{2} \left(\alpha_3^2 - \alpha_2^2 \right) - \frac{1}{2} (\alpha_1 + a) \left(\alpha_3 - \alpha_2 + \frac{\nu}{1-\nu} (1-a) \right) > 0 \tag{30}$$

$$\iff \frac{(\alpha_3 + \alpha_2 - \alpha_1 - a)(\alpha_3 - \alpha_2)}{(\alpha_1 + a)(1-a)} > \frac{\nu}{1-\nu}. \tag{31}$$

Thus, the share of non-reactive consumers must be sufficiently large (low ν).

The comparative statics for the alternative case where the constraint $aQ_r + Q_f \geq x_n$ binds are

$$\begin{pmatrix} \frac{\partial Q_r}{\partial \beta_r} \\ \frac{\partial Q_f}{\partial \beta_r} \end{pmatrix} = -\begin{pmatrix} \frac{1}{\gamma\nu(1-a)|B|}\left(\alpha_3 - \alpha_2 + \frac{\nu}{1-\nu}(1-a)\right) \\ a\frac{\partial Q_r}{\partial \beta_r} + \frac{1}{|B|}\frac{\alpha_3^2 - \alpha_2^2}{2(1-a)\nu\gamma} \end{pmatrix}, \tag{32}$$

where $|B| > 0$ is the determinant of the Hessian matrix for this case (see Appendix C). As for a non-binding constraint, $\frac{\partial Q_r}{\partial \beta_r} < 0$ and the sign of $\frac{\partial Q_f}{\partial \beta_r}$ follows from the relative strength of two countervailing effects. The first term, $-a\frac{\partial Q_r}{\partial \beta_r}$, reflects that fossil capacities are reduced to the extent that the minimal available renewable capacity increases. The second term only occurs in stages *M* and *H*, i.e. if $\alpha_3 - \alpha_2 > 0$ so that case 3 obtains. In this case, renewables are the price-setting technology and, ceteris paribus, higher capacities reduce the price for reactive consumers (see Figure 1). This makes it more attractive to increase supply to non-reactive consumers. Given that $x_n = aQ_r + Q_f$, this must be backed up by additional fossil capacities unless the minimal available renewable capacity increases sufficiently, i.e. unless $a\frac{\partial Q_r}{\partial \beta_r}$ is large. In particular, from (32) we have $\frac{\partial Q_f}{\partial \beta_r} < 0$ if and only if

$$\frac{(\alpha_3 + \alpha_2 - 2a)(\alpha_3 - \alpha_2)}{2a(1-a)} > \frac{\nu}{1-\nu}. \tag{33}$$

The following proposition, which is based on the assumption of a uniform distribution, summarizes these findings.

Proposition 2. *Falling capacity costs of renewables have the following comparative static effects: The efficient level of renewable capacities rises in all diffusion stages. Conversely, the efficient level of fossil capacities falls in diffusion stages* V *and* L *(very low and low renewables), while in stages* M *and* H *(medium and high renewables) this is the case if and only if the share of reactive consumers, ν, is sufficiently high (see conditions (31) and (33) for the case of a non-binding and binding constraint, respectively).*

Having analyzed the two cases where the constraint $aQ_r + Q_f \geq x_n$ binds or not, we now examine what determines which case obtains. For capacity costs just below the level $\overline{\beta_r}$ at which renewables enter the market, αQ_r is very small so that reliable fossils still serve nearly all of the demand of reactive and non-reactive consumers, $x_\nu + x_n$. Thus, we obviously have $Q_f > x_n - aQ_r$ and the comparative statics are as given in (29). Differentiation and substitution for $\frac{\partial Q_f}{\partial \beta_r}$ from (29) yields

$$\frac{\partial x_n}{\partial \beta_r} - a\frac{\partial Q_r}{\partial \beta_r} - \frac{\partial Q_f}{\partial \beta_r} = \frac{1}{2}(\alpha_1 - a)\frac{\partial Q_r}{\partial \beta_r} < 0 \tag{34}$$

due to $\frac{\partial Q_r}{\partial \beta_r} < 0$ and $\alpha_1 > a$. Thus, the left-hand side of $x_n - aQ_r - Q_f \leq 0$ unambiguously increases as β_r falls until the constraint binds; and continues to do so as β_r falls further.[8] Taking into account that we may have entered diffusion stage *R* of renewables only before the constraint starts to bind, we obtain the following corollary.

Corollary 1. *Consider capacity costs of renewables, β_r, that lead to diffusion stage* FR. *Either the constraint $aQ_r + Q_f \geq x_n$ never binds, or there is a threshold level, $\hat{\beta}_r$, such that the constraint does not bind for all $\beta_r > \hat{\beta}_r$, but binds for all $\beta_r < \hat{\beta}_r$.*

[8] Otherwise, there would have to be a β_r for which $h := \frac{\partial x_n}{\partial \beta_r} - a\frac{\partial Q_r}{\partial \beta_r} - \frac{\partial Q_f}{\partial \beta_r} > 0$. However, as long as the constraint binds we obviously have $h = 0$, and for the only other case of a non-binding constraint we have just shown that $h < 0$.

We now take a closer look at how changes in the costs of renewables affect the optimal capacity mix in diffusion stage *FR*, starting with the following example.[9]

Example. $\nu = 0.5$, $A = 100$, $\gamma = 4$, $b_r = 0.1$, $b_f = 6$, $\beta_f = 1$, uniform distribution with $a = 0.01299$.

The black solid and dashed curves in Figure 2 depict efficient levels of renewable and fossil capacities, respectively, as functions of β_r. The vertical dotted lines separate the different diffusion stages, for which central aspects are summarized in the table directly below the figure. We interpret the figure from the right to the left, i.e. for falling capacity costs β_r.

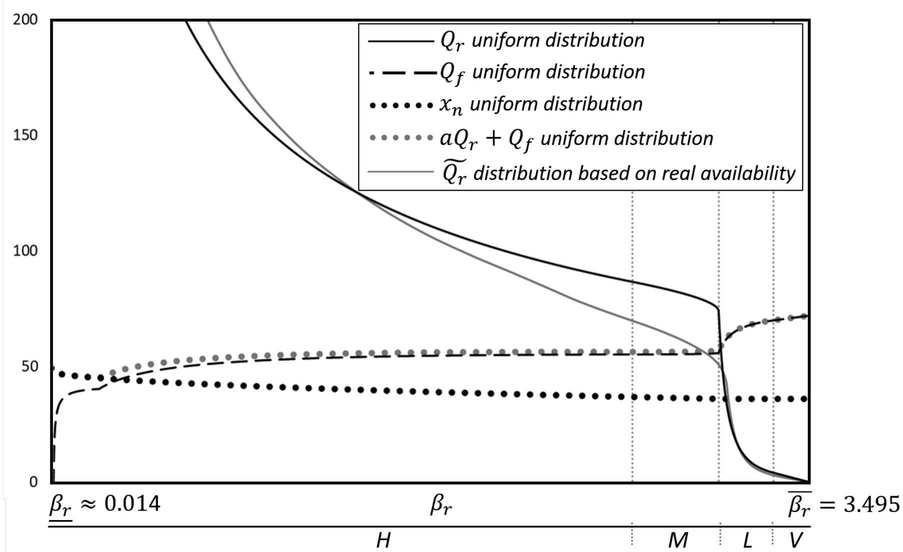
As stated in Lemma 2, it becomes efficient to install renewable capacities once the expected costs of producing one unit of electricity from renewables and fossils are equalized at $\beta_r = \overline{\beta_r}$. Accordingly, as β_r falls below $\overline{\beta_r}$, renewables are cheaper (at full usage) and start to replace fossil capacities until the latter are completely squeezed out of the market at $\beta_r = \underline{\beta_r}$. This pattern is consistent with the preceding analysis. More surprisingly, the diffusion process is not smooth but varies substantially over the different stages.

Despite their cost advantage, renewables substitute the fossil technology only slowly in diffusion stage *V*. The reason is a countervailing effect. Remember that in stage *V* only case 1 obtains, for which the electricity price is falling in the available renewable capacity, αQ_r. Thus, the price is low when a large supply of renewables is available. This reduces their competitiveness compared to fossils that always produce at full capacity in stage *V*. As capacity costs fall further, diffusion stage *L* is reached and the level of renewable capacities increases convexly. This reflects that now case 2 also obtains, in which the price only covers the variable costs of fossils, b_f (see Table 1). Moreover, in contrast to case 1 this price is not falling in the availability of renewables. Both effects improve the relative competitiveness of renewables.

The effect of a reduction in β_r on the efficient capacity level of renewables falls strongly as we reach diffusion stage *M*. The reason is that now also case 3 obtains, for which the price is again falling in the level of renewable capacities, as in case 1. Thus, the price of renewables declines when they have more to sell. The reverse effect applies to fossils. They sell the most when availability of renewables is low and prices are high. Moreover, compared to stage *L,* the maximum price rises.

As we enter diffusion stage *H*, also case 4 obtains, for which the price is constant at marginal production costs of renewables, b_r. Hence, renewables gain no marginal profits for $\alpha > \alpha_3$. Moreover, α_3 is falling in Q_r so that this situation becomes more relevant. However, fossils also suffer because they lie idle more often as renewable capacities rise. This and the low capacity costs of renewables finally dominate so that the marginal effect of a further cost reduction on efficient renewable capacities finally rises until fossils are completely driven out of the market. Note however that this requires extremely low capacity costs of renewables, which is a consequence of the low value for the minimum availability, a, in our specification (and of the lack of storage technologies).

[9] Parameter values have been chosen such that their relation roughly corresponds to real-world data. The value for b_r reflects that renewables have no fuel costs, only some operations and maintenance (O&M) costs in the order of 0 to 3.45 (USD-cent/kWh) for solar PV and 0.25 to 3.47 for wind onshore (IEA 2015). Variable production costs for fossils, c_f, are substantially higher and include fuel, carbon, and variable O&M costs. Fuel costs of natural gas (efficiency of 60%) vary between 3.12 (United States) and 8.19 USD-cent/kWh (OECD Asia); hard coal (efficiency of 46%) costs are around 3.16 USD-cent/kWh (OECD). Carbon costs, δ, vary between 1.01 (natural gas) and 2.21 USD-cent/kWh (hard coal) for a carbon price of 30 USD/tonne CO_2. Variable O&M costs range from 0.27 (USD-cent/kWh) for natural gas-fired plants to 0.34 for coal-fired plants (median values). Depending on the technical lifetime (30 years for natural gas-fired power plants to 40 years for coal-fired power plants), capacity costs of $\beta_f = 1$ (in USD-cent/kWh) correspond to investment costs of 1734 to 1998 USD/kW, which resembles real-world data for a mix of coal-fired and natural gas-fired power plants (all figures are own calculations based on IEA (2015)). The value for A and γ have been chosen so that they reflect an electricity price elasticity of $\varepsilon \approx -0.389$. This reflects a reference price of $p_{ref} = b_f + \beta_f = 7$ with reference load of $L_{ref} = 72$, i.e. we chose diffusion stage F as reference point (see Thimmapuram and Kim (2013), as well as Ito (2014) for further empirical estimates). Finally, the low minimum availability of wind and solar capacities has already been pointed out in footnote 5 (for the exact value see footnote 10 below).



stages	H	M	L	V
cases	1 to 4	1 to 3	1 & 2	1
obtains if	$\alpha_3 < 1$	$\alpha_2 < 1 = \alpha_3$	$\alpha_1 < 1 = \alpha_2$	$\alpha_1 = 1$

Fig. 2: Efficient capacity choices in diffusion stage *FR*

The following proposition, which is based on the assumption of a uniform distribution, generalizes the example. It shows how reductions in the capacity costs of renewables have quite different effects over the various diffusion stages, because the impact of intermittency depends crucially on the market share of renewable capacities.

Proposition 3. *In diffusion stage V, $Q_r^*(\beta_r)$ is a linear function, and in stage* L *it is convex. In stage* M, *$Q_r^*(\beta_r)$ is concave if and only if the share of reactive consumers, ν, is sufficiently high (the conditions are $\frac{\nu}{1-\nu} > \frac{(1-\alpha_2)(1-\alpha_2+\alpha_1-a)}{(2\alpha_2-\alpha_1+a)(1-a)}$ if $aQ_r + Q_f > x_n$, and $\frac{\nu}{1-\nu} > \frac{(1-\alpha_2)(1-\alpha_2+2\alpha_1-2a)}{2(\alpha_2-\alpha_1+a)(1-a)}$ if $aQ_r + Q_f = x_n$). In stage* H, *$Q_r^*(\beta_r)$ is again a convex function.*

Moreover, comparing the related conditions in the Proposition to conditions (31) and (33) that lead to $\frac{\partial Q_f}{\partial \beta_r} < 0$, it is straightforward to show that $\frac{\partial Q_f}{\partial \beta_r} > 0$ is a sufficient condition for $Q_r^*(\beta_r)$ being concave in stage *M*. In line with the assumptions on which Propositions 2 and 3 are based, we have so far considered a uniform distribution for the availability of renewables, α. As a robustness check, the grey solid line in Figure 2 depicts efficient levels of renewable capacities based on real world data for α in Germany in 2015. They roughly correspond to the (truncated) log-normal distribution. In order to focus on the effects that arise from differences in the shape of the distribution, we normalize the real world data such that they have the same reliability $a = 0.01299$ and average availability $\alpha^{ave} = 0.5065$ as in the specification with the uniform distribution.[10] The pattern of capacity levels is similar to that with the uniform distribution, but the kink where renewables are just able to satisfy the whole electricity demand at times of high availability occurs earlier. This is due to the higher maximum availability of the scaled real data distribution.

[10] The real world data yield a minimum availability of 0.00412, a maximum availability of 0.56757, and an average availability of 0.16064 (own calculations based on hourly extrapolated data from the four German transmission system operators, downloaded from www.netztransparenz.de on 22 November 2016). Both values imply that the distribution is substantially less favorable for the diffusion of renewables than the uniform distribution in Example 6. Scaling the real data distribution by the factor 3.1536 yields the same minimum and average availability as for the uniform distribution.

Finally, consider the black dotted curve that depicts non-reactive demand, x_n, and the grey dotted curve that depicts reliable production, $aQ_r + Q_f$ (for the sake of clarity, in the remainder we only show results for the uniform distribution). In line with the discussion before Proposition 2, x_n stays constant in diffusion stages V and L, and rises as β_r falls in diffusion stages M and H. Moreover, the constraint $aQ_r + Q_f \geq x_n$ binds only for very low values of β_r. At the point where it starts to do so, there is a kink in the Q_f curve and the squeeze out of fossil capacities slows down. This reflects that they are now more valuable as a back-up to ensure that supply to non-reactive consumers can always be met.

We now relate the diffusion pattern of renewable energies to the levelized cost of electricity (LCOE), which is the standard metric to assess the competitiveness of different technologies. In our static framework, the LCOE of technology j is the *constant* price for power that equates expected revenues and costs for a representative firm k (see Borenstein 2012). Thus, for $j = r, f$ we have

$$LCOE_j \int_a^1 q_j^k(\alpha)\,dF(\alpha) = b_j \int_a^1 q_j^k(\alpha)\,dF(\alpha) + \beta_j Q_j^k \tag{35}$$

$$\iff LCOE_j = b_j + \frac{\beta_j}{\eta_j}, \tag{36}$$

where $\int_a^1 q_j^k(\alpha)\,dF(\alpha)$ is (expected) output of a firm k that operates with technology j, and $\eta_j := \int_a^1 q_j^k(\alpha)\,dF(\alpha)/Q_j^k$ is its (expected) load factor. Figure 3, which is based on the same parameter specification as Figure 2, depicts the LCOE (black curves, scale on left axis) and the load factor (grey curves, scale on right axis) of renewables (solid) and fossils (dashed) as functions of renewables' capacity costs β_r.

We interpret the figure again in the order of falling capacity costs from the right to the left. First, consider the load factor. Fossil capacities are fully used in diffusion stage V, but their load factor drops by nearly 40 percent in stage L, whereas that of renewables remains constant. This substantially reduces the competitiveness of fossils, and it is the main driver for the sharp rise in renewable capacities in this stage as depicted in Figure 2. Thereafter, i.e. in stages M and H, reductions in the load factors of fossils and renewables differ considerably less. This reflects that the market diffusion is much slower in stages M and H, and mainly driven by the reduction of β_r.

Turning to the LCOE, observe from (36) that $LCOE_r$ increases in β_r and, in addition to this standard effect, the LCOE of both technologies decreases in their respective load factors (remember that b_f, b_r, and β_f are constant). Thus, $LCOE_f$ is increasing as β_r falls because the associated higher share of renewables leads to a lower load factor of fossils (see Figure 3). Obviously, this effect is most pronounced in stage L, where the drop in η_f is largest. By contrast, $LCOE_r$ is decreasing as β_r falls, which shows that decreasing capacity costs dominate the effect of the lower load factor of renewables.

When interpreting $LCOE_j$, it is important to note that Figure 3 depicts its values in the competitive solution, where both technologies are "equally competitive" by construction. Nevertheless, the LCOE of renewables and fossils are equalized only at $\overline{\beta_r}$, the point where renewables just enter the market. For $\beta_r < \overline{\beta_r}$, renewables have lower LCOE. This lends support to the aforementioned criticism of LCOE for use in assessing the competitiveness of renewables, and for the suggestion that the expected market value of renewables should be used instead (see introduction). In our model, the average price or expected "market value" of electricity produced by technology $j = r, f$, denoted p_j, is expected revenues from selling output of technology j to reactive and non-reactive consumers, divided by this output:

$$p_j = \frac{\int_a^{\alpha_j} p_v(\alpha)\left(q_j^k(\alpha) - x_{nj}^k\right) dF(\alpha) + p_n x_{nj}^k \int_a^{\alpha_j} dF(\alpha)}{\int_a^1 q_j^k(\alpha)\,dF(\alpha)}, \tag{37}$$

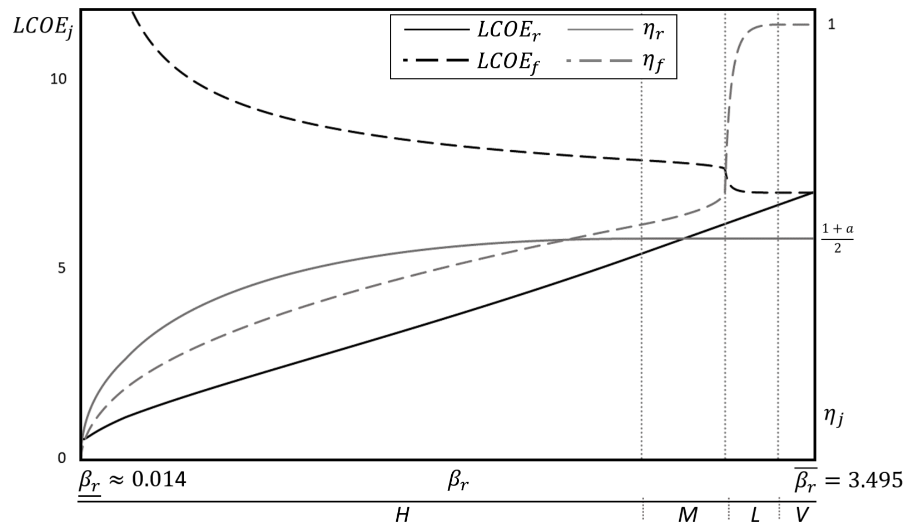


Fig. 3: LCOE and expected load factor in diffusion stage FR

where $\alpha_r = 1$ and $\alpha_f = \alpha_2$. This reflects that fossils produce only for $\alpha \leq \alpha_2$. Moreover, at the competitive solution expected profits of renewable and fossil firms must be zero, i.e. $E\left[\pi_f^k\right] = E\left[\pi_r^k\right] = 0$ (see Section 5). Substituting the expressions for $E\left[\pi_f^k\right]$ and $E\left[\pi_r^k\right]$ as given in (23) and (25), and using this to compare the LCOE defined in (35) with average prices from (37) yields

$$LCOE_r = p_r, \tag{38}$$

$$LCOE_f = p_f + \frac{x_{nf}^k \int_{\alpha_2}^1 (p_n - p_v(\alpha))\, dF(\alpha)}{\int_a^1 q_f^k(\alpha)\, dF(\alpha)}. \tag{39}$$

For renewables, a lower β_r reduces its $LCOE_r$ (see equation (36)). However, $LCOE_r = p_r$ shows that also the market value of renewables falls because their intermittency becomes more obstructive. For the fossil technology we have $LCOE_f > p_f$, i.e. the constant price that fossils need to break even, $LCOE_f$, exceeds the market value of their production, p_f. This is only compatible with zero profits because the market pays fossils also for their reliability to serve the demand of non-reactive consumers. Formally, this is represented by the fraction in (39), where the integral represents the difference between the valuation of production for non-reactive consumers, p_n, and the costs of providing this output, which is $p_v(\alpha) < p_n$ for $\alpha \in (\alpha_2, 1]$.

7 Price caps and the market diffusion of renewables

Lower capacity costs of renewables reduce the costs of producing a given level of electricity. Intuitively, this raises expected electricity production and (weakly) reduces the expected price of reactive consumers in the competitive solution.[11] However, the maximum price does not fall or

[11] The caveat "weakly" accounts for the following consideration. In the efficient solution, there must be no incentives to enter or exit the market. In diffusion stage V, only case 1 obtains, for which fossils always produce at full capacity and sell the same quantity, independent of the price. Therefore, the expected price for reactive consumers p_v must equal their long-run marginal costs, $b_f + \beta_f$. Obviously, the same price obtains in stage F of fossils only. In stage L, case 2 also obtains, for which some fossil capacities lie idle. However, in case 2, the price equals the production

even increases. To see this, remember from Lemma 3 that in diffusion stage FR case 1 always obtains, and that prices are maximal if availability of renewables is at its minimum, i.e. for $\alpha = a$. Using the price for case 1 as given in Table 1, this yields $p_{max} = p_{v1}(a) = \frac{\nu A + x_n - aQ_r - Q_f}{\nu\gamma}$. If $aQ_r + Q_f = x_n$, then $p_{max} = \frac{A}{\gamma}$ and the maximum price is independent of β_r. By contrast, if $aQ_r + Q_f > x_n$, then p_{max} rises as β_r falls—and, thus, renewable capacities rise (see Proposition 3). This is the case because a lower β_r raises the term $x_n - aQ_r - Q_f$ in the expression for p_{max} (see equation (34)).

Regulators may perceive the higher maximum price as politically unacceptable and respond with a price cap; even if it is obvious from the analysis in Section 5 that this is inefficient. In order to be effective, such a price cap, denoted p_c, must be below the maximum price, p_{max}, and above the long-run marginal costs of fossils, $b_f + \beta_f$. Otherwise, fossil capacities would never be built up. This yields $b_f + \beta_f < p_c < \frac{\nu A + x_n - aQ_r - Q_f}{\nu\gamma}$. Moreover, from Table 1 the unregulated variable price exceeds $b_f + \beta_f$ only in case 1. In particular, it equals the price cap if $p_c = p_{v1}(\alpha) = \frac{\nu A + x_n - \alpha Q_r - Q_f}{\nu\gamma}$ and, accordingly, exceeds p_c for all $\alpha < \alpha_c := \frac{\nu(A - \gamma p_c) + x_n - Q_f}{Q_r}$.

Thus, the price cap binds for all $\alpha < \alpha_c$, causing excess demand of reactive consumers (non-reactive consumers are unaffected). It is well known that this leads to welfare losses when consumers are served randomly.[12] We abstract from this complication as it is not the focus of our article. Instead, we assume (as in Joskow and Tirole 2007) that consumers are served according to their willingness to pay, as it would be the case without a price cap. For given capacity levels, therefore, the price cap would not affect total welfare, but lead to a shift of (expected) surplus from producers to consumers because the latter pay a lower price if the cap binds.

This, however, affects firms' incentives to invest in electricity production capacities. First, consider the case of $aQ_r + Q_f > x_n$ for which both, renewable and fossil firms, serve reactive consumers. Fossils sell most of their output to reactive consumers when prices are high and, therefore, in situations where the price cap binds. For renewables the opposite pattern applies so that less of their output is affected by the price cap. Accordingly, a price cap raises the competitiveness of renewables as compared to fossils. The opposite result obtains for $aQ_r + Q_f = x_n$. The reason is simply that fossil firms do not sell to reactive consumers in this case (see Section 5). Therefore, only renewables obtain lower revenues if the price cap binds and their competitiveness is reduced.[13] We obtain the following results.

Proposition 4. *Consider a competitive electricity market with fossil and renewable capacities. As capacity costs of renewables fall, the unregulated maximum price of reactive consumers remains unchanged if constraint $aQ_r + Q_f \geq x_n$ binds, but increases if it is non-binding. Implementation of a price cap $p_c \in \left(b_f + \beta_f, \frac{\nu A + x_n - aQ_r - Q_f}{\nu\gamma}\right)$ has the following effects (assuming a uniform distribution and denoting the equilibrium by superscript c):*

1. *If $aQ_r^c + Q_f^c > x_n^c$, then renewable capacities are inefficiently high, i.e. $Q_r^c > Q_r^*$. Fossil capacities are inefficiently low in stages* V *and* L, *i.e. $Q_f^c < Q_f^*$, but in stages* M *and* H *the effect is inconclusive.*

2. *If $aQ_r^c + Q_f^c = x_n^c$, then the above effects are reversed, i.e. $Q_r^c < Q_r^*$ and , moreover, in stages* V *and* L we have *$Q_f^c > Q_f^*$, while in stages* M *and* H *the effect is inconclusive.*

costs of fossils, b_f. Hence expected profits of fossils would remain unchanged if they sold their entire capacity at this price. Thus, the expected price must again be equal to the long-run marginal costs of fossils. In conclusion, $E\left[p_v^F\right] = E\left[p_v^V\right] = E\left[p_v^L\right] = b_f + \beta_f$, where superscripts represent the diffusion stages and E is the expectation operator. Note that this is a standard result in the peak load pricing literature (e.g. Green and Léautier (2017, Lemma 1)).

[12] See Visscher (1973) for a seminal contribution on different rationing schemes as well as Crew et al. (1995) for a survey of different ways of interpreting rationing.

[13] Note that these mechanisms do not depend on the (admittedly debatable) assumption of efficient rationing.

8 Concluding Remarks

Countries with long-term renewable energy targets typically plan that the market share of renewables increases roughly linear over time. For example, in Germany the planned shares of primary and final energy from renewable sources are 18% by 2020, 30% by 2030, 45% by 2040 and 60% by 2050 (REN21 2017, p. 187). In this paper, we have analyzed the efficient market diffusion of intermittent renewable energies as their capacity costs fall over time. We have found an S-shaped pattern, and showed how this can be traced back to the effects of intermittency. In particular, it becomes efficient to add renewables to the electricity generation system when their LCOE has fallen to that of fossils. Initially, further cost reductions have only a small effect on their efficient market share, but this effect increases substantially as soon as renewables reduce the load factor of fossils. Once the level of renewable capacities is large enough to satisfy the entire energy demand at times of high availability, the marginal effect of a further cost reduction on the efficient level of renewable capacities decreases again.

Obviously, the real world diffusion of renewable energies is heavily influenced by policies designed to promote them. This is particularly the case in countries that use a feed-in tariff (FIT) because it shelters renewables from low prices at times of high availability. By contrast, apart from the price cap in Section 7, the only policy instrument in our model is a tax that internalizes the environmental costs of fossil fuels. Nevertheless, consistent with our model, in many countries fossil energies suffer from a reduced load and from low prices at times of high availability of renewables (Cludius, Hermann, Matthes, and Graichen, 2014). This has strengthened the relative competitiveness of renewables and contributed to their sharp rise that we observe in the real world as well as in our model.

Currently, many countries intend to curtail the subsidization of renewables while at the same time proclaiming very ambitious targets for the expansion of renewable energies. For example, in Germany the share of RES in the electricity grid is targeted to rise from the current level of 33 per cent to 80-95 per cent in 2050. This seems to be driven by the perception that reductions in the capacity costs of renewables can compensate the lower subsidies, thereby upholding high growth rates for renewables. Our analysis suggests that this neglects the impediments to the diffusion of renewables that result from the substantial drop in their market value, once they are able to satisfy the whole energy demand at times of maximum availability.[14] Thus, if the target is an energy system that is completely based on renewables, then the most difficult stages of the energy transition may still lie ahead, even if capacity costs of renewables continue to fall rapidly.

High shares of renewables in the energy mix lead to extended periods in which wholesale electricity prices are essentially zero. Our analysis suggests that the efficient outcome can nonetheless be implemented by competitive markets, provided that environmental costs of fossil fuels are internalized by a Pigouvian tax—and abstracting from other "technical" complications such as grid-related requirements (see Abrell and Rausch 2016) and ramping costs. This is the case even if only a subgroup of consumers can adapt its demand to electricity prices that fluctuate with the availability of renewables. Nevertheless, a higher share of reactive consumers obviously raises welfare, as this would allow to better align demand to scarcities as expressed by prices. It would also eliminate the somewhat surprising finding that lower costs of renewables may raise the efficient level of fossils.

However, confronting more consumers with dynamic electricity prices implies a stronger exposure to high maximum prices that may obtain in a largely renewables-based energy system. Price caps may help to cushion unwanted social effects of this, but they affect investment incentives. Our analysis suggests that these effects change as the level of renewable capacities rises. Initially,

[14] In 2015, average load in Germany was 57.55 GW, maximum supply of renewables 47.63 GW, and the lowest residual load 6.73 GW. Renewable production is highest between 12 am and 4 pm. Average load during these hours is 63.1 GW but minimum load only 42.5 GW (own compilations based on data from the four German transmission system operators, downloaded from www.netztransparenz.de and www.entsoe.de on 11 January 2016). Hence, theoretically, the installed capacity is already capable of satisfying the total demand at times of high supply and low demand.

fossil production (the peak technology) suffers most from a price cap. Later, fossils are mainly used as a backup to ensure supply to non-reactive consumers so that the burden of a price cap falls upon renewables.

An obvious extension of the above analysis would be the integration of dynamic aspects of investment decisions. Moreover, we have not considered measures that are targeted at dampening the effects of intermittency. Storage and international trade in electricity that exploits regional differences in the availability of intermittent renewables are two candidates. Presumably, this and other changes to make the model more general—e.g. the integration of more than two technologies—would require greater reliance on numerical simulations.

References

Abrell, J. and S. Rausch (2016). Cross-country electricity trade, renewable energy and European transmission infrastructure policy. *Journal of Environmental Economics and Management 79*, 87–113.

Abrell, J., S. Rausch, and C. Streitberger (2018). The economics of renewable energy support. Working Paper 18/286, CER-ETH Zurich.

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

Ambec, S. and C. Crampes (2017). Decarbonizing electricity generation with intermittent sources of energy. *TSE Working Paper n. 15-603, revised version.*

Andor, M. and A. Voss (2016). Optimal renewable-energy promotion: Capacity subsidies vs. generation subsidies. *Resource and Energy Economics 45*, 144–158.

Borenstein, S. (2012). The private and public economics of renewable electricity generation. *Journal of Economic Perspectives 26*(1), 67–92.

Borenstein, S. and S. Holland (2005). On the efficiency of competitive electricity markets with time-invariant retail prices. *The Rand Journal of Economics 36*(3), 469–93.

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

Cludius, J., H. Hermann, F. C. Matthes, and V. Graichen (2014). The merit order effect of wind and photovoltaic electricity generation in Germany 2008-2016: Estimation and distributional implications. *Energy Economics 44*, 302–313.

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.

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.

Faruqui, A. and S. Sergici (2010). Household response to dynamic pricing of electricity: A survey of 15 experiments. *Journal of Regulatory Economics 38*(2), 193–225.

Fell, H. and J. Linn (2013). Renewable electricity policies, heterogeneity, and cost effectiveness. *Journal of Environmental Economics and Management 66*(3), 688 – 707.

Fischer, C., L. Preonas, and R. G. Newell (2017). Environmental and technology policy options in the electricity sector: Are we deploying too many? *Journal of the Association of Environmental and Resource Economists 4*(4), 959–984.

Garcia, A., J. M. Alzate, and J. Barrera (2012). Regulatory design and incentives for renewable energy. *Journal of Regulatory Economics 41*(3), 315–336.

Geroski, P. A. (2000). Models of technology diffusion. *Research Policy 29*(4), 603–625.

Gowrisankaran, G., S. S. Reynolds, and M. Samano (2016). Intermittency and the value of renewable energy. *Journal of Political Economy 124*(4), 1187–1234.

Green, R. and N. Vasilakos (2010). Market behaviour with large amounts of intermittent generation. *Energy Policy 38*(7), 3211–3220.

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.*

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica 25*(4), 501–522.

Hirth, L. (2013). The market value of variable renewables: The effect of solar wind power variability on their relative price. *Energy Economics 38*, 218–236.

IEA (2015). Projected costs of generating electricity 2015 edition. Technical report, International Energy Agency.

IRENA (2015). Rethinking energy 2015. Technical report, International Renewable Energy Agency.

IRENA (2016). The power to change: Solar and wind cost reduction potential to 2025. Technical report, International Renewable Energy Agency.

Ito, K. (2014). Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *American Economic Review 104*(2), 537–563.

Iversen, E. B., J. M. Morales, J. K. Møller, and H. Madsen (2016). Short-term probabilistic forecasting of wind speed using stochastic differential equations. *International Journal of Forecasting 32*(3), 981–990.

Jessoe, K. and D. Rapson (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review 104*(4), 1417–1438.

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

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

Joskow, P. L. and C. D. Wolfram (2012). Dynamic pricing of electricity. *American Economic Review 102*(3), 381–385.

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

Liski, M. and I. Vehviläinen (2016). Gone with the wind? An empirical analysis of the renewable energy rent transfer. Cambridge Working Paper in Economics 1701, University of Cambridge.

REN21 (2017). Renewables 2017 global status report. Technical report, Paris.

Requate, T. and W. Unold (2003). Environmental policy incentives to adopt advanced abatement technology: Will the true ranking please stand up? *European Economic Review 47*(1), 125–146.

Rouillon, S. (2015). Optimal and equilibrium investment in the intermittent generation technologies. *Revue d'économie politique 125*(3), 415–452.

Simon, C. P. and L. Blume (1994). *Mathematics for economists.* Norton New York.

Stoft, S. (2003). The demand for operating reserves: Key to price spikes and investment. *Power Systems, IEEE Transactions on 18*(2), 470–477.

Sydsaeter, K., P. Hammond, A. Seierstad, and A. Strom (2005). *Further mathematics for economic analysis.* Pearson Education.

Thimmapuram, P. R. and J. Kim (2013). Consumers' price elasticity of demand modeling with economic effects on electricity markets using an agent-based model. *IEEE Transactions on Smart Grid 4*(1), 390–397.

Twomey, P. and K. Neuhoff (2010). Wind power and market power in competitive markets. *Energy Policy 38*(7), 3198–3210.

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

Appendix

A Proof of Lemma 1

Diffusion stage *FR*. First, consider the possibility that $x_v^* = 0$. Using the specification of inverse demand and (7), this implies $p_v = \frac{A}{\gamma} \leq \lambda$. By assumption, $\frac{A}{\gamma} > b_f > b_r$ so that from (8) and (9) we have $q_r = \alpha Q_r$ and $q_f = Q_f$. However, demand of non-reactive consumers satisfies $x_n \leq aQ_r + Q_f \leq \alpha Q_r + Q_f = q_r + q_f$, with equality if and only if $\alpha = a$. This is a measure zero event that can be ignored in the remainder. For all $\alpha > a$, $x_v^* = 0$ can not be an efficient solution as it would imply $q_r + q_f > x_n + x_v$, i.e. supply would exceed demand.

For $x_v^* > 0$, there are different combinations of binding and non-binding capacity constraints: First, we may have $\mu_r(\alpha), \mu_f(\alpha) > 0$. From the complementary slackness conditions (10) and (11) it follows that $q_r(\alpha) = \alpha Q_r$ and $q_f(\alpha) = Q_f$ (case 1). Second, $\mu_r(\alpha) > \mu_f(\alpha) = 0$ implies $q_r(\alpha) = \alpha Q_r$ and $q_f(\alpha) \leq Q_f$. We must distinguish whether (9) binds or not. If it does, we have $q_f(\alpha) > 0$, so that (using 7) $b_f = \lambda = p_v$ and (using inverse demand) $x_v = \nu(A - \gamma b_f)$ as well as $q_f(\alpha) = x_v + x_n - \alpha Q_r$ (case 2). Alternatively, $q_f(\alpha) = 0$ for which $b_f > \lambda = \frac{\nu A - (\alpha Q_r - x_n)}{\nu\gamma} = p_v$ (case 3), where we have used (7) and $x_\nu = \alpha Q_r - x_n$ to obtain the price. Third, we may have $\mu_r(\alpha) = \mu_f(\alpha) = 0$ for which (8) and (9) cannot bind simultaneously. In particular, because $b_r < b_f$ only (8) binds and $q_f(\alpha) = 0$. Moreover, (7) and the binding condition (8) imply that $p_\nu = b_r$. From the inverse demand function, we then obtain $x_\nu = \nu(A - \gamma b_r)$ and $q_r = \nu(A - \gamma b_r) + x_n$ (case 4).[15]

The threshold value α_1 that separates cases 1 and 2 is characterized by $\mu_f(\alpha) = 0$ and $q_f(\alpha) + q_r(\alpha) = \alpha Q_r + Q_f$. From (7) and (9), this yields $p_v = \frac{\nu A - (\alpha_1 Q_r + Q_f - x_n)}{\nu\gamma} = b_f$. The threshold value α_2 that separates cases 2 and 3 is characterized by $p_v = \frac{\nu A - (\alpha_2 Q_r - x_n)}{\nu\gamma} = b_f$. Finally, the threshold value α_3 that separates cases 3 and 4 is characterized by $p_v = \frac{\nu A - (\alpha_3 Q_r - x_n)}{\nu\gamma} = b_r$.

Diffusion stage *R*. If there are no fossil capacities, the choice variable q_f can be dropped from the maximization problem so that the solution follows from (7), (8) and (10). Condition (8) always binds (by the assumption that the maximum WTP exceeds the variable cost of renewables) so that together with (7) we obtain $p_\nu(\alpha) = b_r + \mu_r(\alpha)$. From the complementary slackness condition (10) there are two cases. First, $\mu_r(\alpha) > 0$ and $q_r(\alpha) = \alpha Q_r$ so that $x_v = \alpha Q_r - x_n$ and $p_\nu = \frac{\nu A - (\alpha Q_r - x_n)}{\nu\gamma} > b_r$. Second, $\mu_r(\alpha) = 0$ and $q_r(\alpha) \leq \alpha Q_r$ so that $p_\nu = b_r$, $x_v = \nu(A - \gamma b_r)$ and $q_r(\alpha) = \nu(A - \gamma b_r) + x_n$. Finally, $\mu_r(\alpha) = \frac{\nu A - x_\nu(\alpha)}{\nu\gamma} - b_r \geq 0$ would be violated for all $x_\nu(\alpha) = \alpha Q_r - x_n > \nu(A - \gamma b_r)$, which defines the threshold value $\alpha_3 = \frac{\nu(A - \gamma b_r) + x_n}{Q_r}$.

B Proof of Lemma 3

In contradiction to the first statement, suppose that we are in diffusion stage FR and $a \geq \alpha_1$. Hence case 1 never obtains, and $p_{\nu 2} = b_f$ in case 2. Accordingly, $\int_a^{\alpha_2} (p_v(\alpha) - b_f)\, dF(\alpha) = 0$ and the first-order condition (19) becomes $-\beta_f + \zeta = 0$. However, in the efficient solution, the cost of an additional unit of fossil capacities, β_f, is obviously higher than the shadow price ζ of relaxing the constraint $aQ_r + Q_f \geq x_n$ by this additional unit (otherwise, this additional unit would have been installed). Hence $-\beta_f + \zeta < 0$, a contradiction. Turning to the second statement, suppose by contradiction that we are in diffusion stage R and $a \geq \alpha_3$. Hence only case 4 would obtain for which $p_\nu = b_r$, and it cannot be efficient to incur the costs of installing renewable capacities.

[15] A priori, there is a further combination of multipliers, namely $\mu_f(\alpha) > \mu_r(\alpha) = 0$, for which $q_f(\alpha) = Q_f$ by (11). Thus (9) binds, which yields a contradiction because the left-hand side of (8) is always larger than the left-hand side of (9) for $\mu_f(\alpha) > \mu_r(\alpha)$. Hence, this case cannot occur.

C Proof of Proposition 2

It remains to determine the comparative statics in (29) and (32). The proposition relates to diffusion stage *FR*, for which $Q_r, Q_f, x_n > 0$. First, consider the case of $aQ_r + Q_f > x_n$ so that $\zeta = 0$ by complementary slackness. Thus, efficient capacity levels follow from the system of the three (binding) first-order conditions (19) to (21) for $\zeta = 0$. For parsimony, denote the second-order derivatives by W_{ij}; $i, j = r, f, n$, where the two subscripts represent the variables of differentiation, Q_r, Q_f, and x_n (e.g. $W_{rf} := \frac{\partial W_{Q_r}}{\partial Q_f}$). Applying the implicit function theorem (see Simon and Blume 1994, p. 355), thereby using $\frac{\partial W_{Q_r}}{\partial \beta_r} = -1$ and $\frac{\partial W_{Q_f}}{\partial \beta_r} = \frac{\partial W_{x_n}}{\partial \beta_r} = 0$ yields

$$\begin{aligned} \begin{pmatrix} \frac{\partial Q_r}{\partial \beta_r} \\ \frac{\partial Q_f}{\partial \beta_r} \\ \frac{\partial x_n}{\partial \beta_r} \end{pmatrix} &= -\begin{pmatrix} W_{rr} & W_{rf} & W_{rn} \\ W_{fr} & W_{ff} & W_{fn} \\ W_{nr} & W_{nf} & W_{nn} \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} && (40) \\ &= \frac{1}{|A|} \begin{pmatrix} W_{ff}W_{nn} - W_{fn}W_{nf} \\ W_{fn}W_{nr} - W_{fr}W_{nn} \\ W_{fr}W_{nf} - W_{ff}W_{nr} \end{pmatrix}. && (41) \end{aligned}$$

Here, $|A|$ denotes the determinant of the Hessian matrix A in (40), and we have taken into account that all terms other than the first column in A^{-1} are multiplied by 0 and, therefore, cancel. Using prices as given in Table 1, for a uniform distribution the second-order derivatives can be calculated from (19) to (21) as (remember that the first-order derivatives are continuous so that the effects at the borders of the integrals are canceling out)

$$W_{fr} = W_{rf} = -\int_a^{\alpha_1} \frac{\alpha}{\nu\gamma} dF(\alpha) = -\frac{\alpha_1^2 - a^2}{2\nu\gamma(1-a)} < 0, \tag{42}$$

$$W_{fn} = W_{nf} = -W_{ff} = \int_a^{\alpha_1} \frac{1}{\nu\gamma} dF(\alpha) = \frac{\alpha_1 - a}{\nu\gamma(1-a)} > 0, \tag{43}$$

$$W_{rn} = W_{nr} = \int_a^{\alpha_1} \frac{\alpha}{\nu\gamma} dF(\alpha) + \int_{\alpha_2}^{\alpha_3} \frac{\alpha}{\nu\gamma} dF(\alpha) = \frac{\alpha_3^2 - \alpha_2^2 + \alpha_1^2 - a^2}{2\nu\gamma(1-a)} > 0, \tag{44}$$

$$W_{rr} = -\int_a^{\alpha_1} \frac{\alpha^2}{\nu\gamma} dF(\alpha) - \int_{\alpha_2}^{\alpha_3} \frac{\alpha^2}{\nu\gamma} dF(\alpha) = -\frac{\alpha_3^3 - \alpha_2^3 + \alpha_1^3 - a^3}{3\nu\gamma(1-a)} < 0, \tag{45}$$

$$W_{nn} = -\frac{1}{(1-\nu)\gamma} - \int_a^{\alpha_1} \frac{1}{\nu\gamma} dF(\alpha) - \int_{\alpha_2}^{\alpha_3} \frac{1}{\nu\gamma} dF(\alpha) = -\frac{1}{(1-\nu)\gamma} - \frac{\alpha_3 - \alpha_2 + \alpha_1 - a}{\nu\gamma(1-a)} < 0. \tag{46}$$

Applying Sarrus rule, substituting from above and rearranging, the determinant becomes

$$\begin{aligned} |A| &= W_{rr}W_{ff}W_{nn} + W_{rf}W_{fn}W_{nr} + W_{rn}W_{fr}W_{nf} - W_{rn}^2 W_{ff} - W_{rr}W_{fn}^2 - W_{rf}^2 W_{nn} \\ &= (a - \alpha_1) \frac{(\alpha_3 - \alpha_2)^4 + (\alpha_3 - \alpha_2)(\alpha_1 - a)^3 + \frac{\nu(1-a)}{1-\nu}\left[(\alpha_1 - a)^3 + 4\left(\alpha_3^3 - \alpha_2^3\right)\right]}{12\nu^3\gamma^3(1-a)^3}, \end{aligned} \tag{47}$$

which is negative. In addition, observe that from the above we have $W_{rr} < 0$ and

$$\begin{vmatrix} W_{rr} & W_{rf} \\ W_{fr} & W_{ff} \end{vmatrix} = (\alpha_1 - a) \frac{(\alpha_1 - a)^3 + 4\left(\alpha_3^3 - \alpha_2^3\right)}{12\nu^2\gamma^2(1-a)^2} > 0.$$

Therefore, the Hessian matrix A is negative definite so that $W(\mathbf{Q})$ is concave as claimed in Section 4 regarding the second-order conditions. Moreover, substitution of the second-order derivatives into (41) and rearranging yields expression (29) in the main text.

We now turn to the analysis of the second case of $aQ_r+Q_f = x_n$ so that $\zeta \geq 0$ by complementary slackness. Solving (21) for ζ, substituting this into (19) and (20), and using $x_n = aQ_r + Q_f$, the first-order conditions become

$$W_{Q_f} = -\int_a^{\alpha_2} b_f dF(\alpha) - \int_{\alpha_2}^1 p_v(\alpha)\, dF(\alpha) + \frac{(1-\nu)A - aQ_r - Q_f}{(1-\nu)\gamma} - \beta_f = 0, \tag{48}$$

$$W_{Q_r} = \int_a^1 (p_v(\alpha)(\alpha - a) - b_r\alpha)\, dF(\alpha) + a\frac{(1-\nu)A - aQ_r - Q_f}{(1-\nu)\gamma} - \beta_r = 0. \tag{49}$$

Applying the implicit function theorem yields (using $\partial W_{Q_f}/\partial\beta_r = 0$ and $\partial W_{Q_r}/\partial\beta_r = -1$)

$$\begin{pmatrix} \frac{\partial Q_r}{\partial \beta_r} \\ \frac{\partial Q_f}{\partial \beta_r} \end{pmatrix} = \frac{1}{|B|}\begin{pmatrix} W_{ff} \\ -W_{fr} \end{pmatrix}, \tag{50}$$

where $|B| := W_{rr}W_{ff} - W_{rf}W_{fr}$ denotes the determinant of the Hessian matrix for this case and the partial derivatives on the right-hand side follow from differentiation of the first-order conditions (48) and (49). Upon substitution of $x_n = aQ_r+Q_f$ in Table 1, we obtain $\alpha_1 = \frac{\nu(A-\gamma b_f)+aQ_r}{Q_r}$, $\alpha_2 = \frac{\nu(A-\gamma b_f)+aQ_r+Q_f}{Q_r}$, $\alpha_3 = \frac{\nu(A-\gamma b_r)+aQ_r+Q_f}{Q_r}$ as well as prices $p_{v1}(\alpha) = \frac{\nu A-(\alpha-a)Q_r}{\nu\gamma}$ and $p_{v3}(\alpha) = \frac{\nu A-(\alpha-a)Q_r+Q_f}{\nu\gamma}$. Using this, for a uniform distribution the second-order derivatives can be calculated from (48) and (49) as (as above, the effects over the borders of the integrals are canceling out because the first-order derivatives are continuous)

$$W_{ff} = -\int_{\alpha_2}^{\alpha_3} \frac{1}{\nu\gamma} dF(\alpha) - \frac{1}{(1-\nu)\gamma} = -\frac{\alpha_3 - \alpha_2}{(1-a)\nu\gamma} - \frac{1}{(1-\nu)\gamma} < 0, \tag{51}$$

$$W_{fr} = W_{rf} = \int_{\alpha_2}^{\alpha_3} \frac{\alpha - a}{\nu\gamma} dF(\alpha) - \frac{a}{(1-\nu)\gamma} = aW_{ff} + \frac{\alpha_3^2 - \alpha_2^2}{2(1-a)\nu\gamma}, \tag{52}$$

$$\begin{aligned} W_{rr} &= -\int_a^{\alpha_1} \frac{(\alpha-a)^2}{\nu\gamma} dF(\alpha) - \int_{\alpha_2}^{\alpha_3} \frac{(\alpha-a)^2}{\nu\gamma} dF(\alpha) - \frac{a^2}{(1-\nu)\gamma} \\ &= -\frac{(\alpha_1 - a)^3}{3(1-a)\nu\gamma} - \frac{(\alpha_3-a)^3 - (\alpha_2-a)^3}{3(1-a)\nu\gamma} - \frac{a^2}{(1-\nu)\gamma} < 0. \end{aligned} \tag{53}$$

Upon substitution and some rearrangements, the determinant of the Hessian becomes

$$|B| = \frac{(\alpha_3-\alpha_2)^4 + 4(\alpha_3-\alpha_2)(\alpha_1-a)^3 + \frac{\nu(1-a)}{(1-\nu)}\left[4(\alpha_1-a)^3 + 4\left(\alpha_3^3 - \alpha_2^3\right)\right]}{12\nu^2\gamma^2(1-a)^2}, \tag{54}$$

which is clearly positive. Together with $W_{rr} < 0$ this shows that also in the case of $aQ_r + Q_f = x_n$ the Hessian matrix is negative definite and $W(\mathbf{Q})$ is concave (see Section 4). Substitution into (50) and rearranging yields (32).

D Proof of Proposition 3

In order to prove the statements regarding convexity and concavity, we need to analyze the second-order derivatives of Q_r w.r.t. β_r. As in the proof of Proposition 2, we first analyse the case $aQ_r + Q_f > x_n$ for which the comparative statics are given by (29). In diffusion stage V, $\alpha_1 = 1$ so that the expression for $\frac{\partial Q_r}{\partial \beta_r}$ contains only exogenous parameters and is independent of β_r; hence $\frac{\partial^2 Q_r}{\partial \beta_r^2} = 0$. The results regarding the other diffusion stages require an explicit calculation of $\frac{\partial^2 Q_r}{\partial \beta_r^2}$ as $\alpha_1 = min\{(\nu A - \nu\gamma b_f - Q_f + x_n)/Q_r, 1\}, \alpha_2 = min\{(\nu(A - \gamma b_f) + x_n)/Q_r, 1\}$ and $\alpha_3 = min\{(\nu(A - \gamma b_r) + x_n)/Q_r, 1\}$ are functions of Q_r, Q_f, x_n and, thus, depend on β_r. Doing so yields (after some tedious transformations)

$$\frac{\partial^2 Q_r}{\partial \beta_r^2}\frac{\kappa^2}{36\nu\gamma(1-a)} = \frac{\partial \alpha_3}{\partial \beta_r}\left[(\alpha_3-\alpha_2)^2 + 2\alpha_3\frac{\nu(1-a)}{1-\nu}\right]^2 - \frac{\partial \alpha_2}{\partial \beta_r}\left[(\alpha_3-\alpha_2)^2 - 2\alpha_2\frac{\nu(1-a)}{1-\nu}\right]^2 + \frac{\partial \alpha_1}{\partial \beta_r}(\alpha_1 - a)^2\left[(\alpha_3-\alpha_2) + \frac{\nu(1-a)}{1-\nu}\right]^2, \tag{55}$$

where

$$\kappa := \frac{12\nu^3\gamma^3(1-a)^3|A|}{a-\alpha_1} > 0.$$

In diffusion stage *L*, $\alpha_1 < \alpha_2 = \alpha_3 = 1$ so that $\frac{\partial \alpha_2}{\partial \beta_r} = \frac{\partial \alpha_3}{\partial \beta_r} = 0$. Moreover, differentiation of α_1, thereby using (29) and the definition of κ, yields

$$\frac{\partial \alpha_1}{\partial \beta_r} = \frac{6\nu\gamma(1-a)}{\kappa Q_r}(\alpha_1 - a)\left[(\alpha_3-\alpha_2) + \frac{\nu(1-a)}{1-\nu}\right] > 0. \tag{56}$$

Accordingly, the terms in the first line of (55) cancel and it follows straightforwardly that $\frac{\partial^2 Q_r}{\partial \beta_r^2} > 0$, proving convexity of $Q_r(\beta_r)$ in stage *L*. In diffusion stage *M*, $\alpha_1 < \alpha_2 < \alpha_3 = 1$ so that $\frac{\partial \alpha_3}{\partial \beta_r} = 0$, while differentiation of α_2 yields

$$\frac{\partial \alpha_2}{\partial \beta_r} = -\frac{6\nu\gamma(1-a)}{\kappa Q_r}\left[(\alpha_3-\alpha_2)^2 - 2\alpha_2\frac{\nu(1-a)}{1-\nu}\right]. \tag{57}$$

Substitution of this and (56) into (55) yields

$$\frac{\partial^2 Q_r}{\partial \beta_r^2}\frac{\kappa^3 Q_r}{216[\nu\gamma(1-a)]^2} = \left[(1-\alpha_2)^2 - 2\alpha_2\frac{\nu(1-a)}{1-\nu}\right]^3 + (\alpha_1-a)^3\left[(1-\alpha_2) + \frac{\nu(1-a)}{1-\nu}\right]^2.$$

This term is negative and, thus, $Q_r(\beta_r)$ is concave in stage *M*, if and only if

$$\frac{(1-\alpha_2)(1-\alpha_2+\alpha_1-a)}{(2\alpha_2-\alpha_1+a)(1-a)} < \frac{\nu}{1-\nu}.$$

Finally, in diffusion stage *H*, $\alpha_1 < \alpha_2 < \alpha_3 < 1$, where

$$\frac{\partial \alpha_3}{\partial \beta_r} = \frac{6\nu\gamma(1-a)}{\kappa Q_r}\left[(\alpha_3-\alpha_2)^2 + 2\alpha_3\frac{\nu(1-a)}{1-\nu}\right] > 0. \tag{58}$$

Substitution of this, (56) and (57) into (55) yields

$$\frac{\partial^2 Q_r}{\partial \beta_r^2}\frac{\kappa^3 Q_r}{216[\nu\gamma(1-a)]^2} = \left[(\alpha_3-\alpha_2)^2 + 2\alpha_3\frac{\nu(1-a)}{1-\nu}\right]^3 + \left[(\alpha_3-\alpha_2)^2 - 2\alpha_2\frac{\nu(1-a)}{1-\nu}\right]^3 + (\alpha_1-a)^3\left[(\alpha_3-\alpha_2) + \frac{\nu(1-a)}{1-\nu}\right]^3.$$

Noting that $2\alpha_3 > 2\alpha_2$, this term is strictly positive and, thus, $Q_r(\beta_r)$ is convex in stage *H*.

The proof for the alternative case of $aQ_r + Q_f = x_n$ proceeds along similar lines. The comparative statics are now given by (32). Differentiation (and some tedious transformations) yield

$$\frac{\partial^2 Q_r}{\partial \beta_r^2}\frac{z^2}{36\nu\gamma(1-a)} = \frac{\partial \alpha_3}{\partial \beta_r}\left[(\alpha_3-\alpha_2)^2 + 2\alpha_3\frac{\nu(1-a)}{1-\nu}\right]^2 - \frac{\partial \alpha_2}{\partial \beta_r}\left[(\alpha_3-\alpha_2)^2 - 2\alpha_2\frac{\nu(1-a)}{1-\nu}\right]^2 + \frac{\partial \alpha_1}{\partial \beta_r}4(\alpha_1-a)^2\left[(\alpha_3-\alpha_2)+\frac{\nu(1-a)}{1-\nu}\right]^2, \quad (59)$$

where

$$z := 12\nu^2\gamma^2(1-a)^2|B| > 0.$$

In diffusion stage *V*, $\alpha_1 = \alpha_2 = \alpha_3 = 1$ so that $\frac{\partial^2 Q_r}{\partial \beta_r^2} = 0$; hence $Q_r(\beta_r)$ is linear. For the other cases, we again need to determine $\frac{\partial \alpha_1}{\partial \beta_r}, \frac{\partial \alpha_2}{\partial \beta_r}, \frac{\partial \alpha_3}{\partial \beta_r}$, using now the comparative statics in (32), rather than those in (29) that we used in the first part of the proof. Doing so yields

$$\frac{\partial \alpha_1}{\partial \beta_r} = \frac{6\nu\gamma(1-a)}{zQ_r}2(\alpha_1-a)\left[(\alpha_3-\alpha_2)+\frac{\nu(1-a)}{1-\nu}\right] > 0, \quad (60)$$

whereas $\frac{\partial \alpha_2}{\partial \beta_r}$ and $\frac{\partial \alpha_3}{\partial \beta_r}$ are the same as the derivatives (57) and (58) if one replaces κ by z. In diffusion stage *L*, $\alpha_1 < \alpha_2 = \alpha_3 = 1$ and $\frac{\partial \alpha_1}{\partial \beta_r} > 0$. From (59) it follows immediately that $\frac{\partial^2 Q_r}{\partial \beta_r^2} > 0$, hence $Q_r(\beta_r)$ is convex in stage *L*. In diffusion stage *M*, $\alpha_1 < \alpha_2 < \alpha_3 = 1$ so that upon substitution of $\frac{\partial \alpha_1}{\partial \beta_r}, \frac{\partial \alpha_2}{\partial \beta_r}$ into (59) yields

$$\frac{\partial^2 Q_r}{\partial \beta_r^2}\frac{z^3 Q_r}{216[\nu\gamma(1-a)]^2} = \left[(1-\alpha_2)^2 - 2\alpha_2\frac{\nu(1-a)}{1-\nu}\right]^3 + 8(\alpha_1-a)^3\left[(1-\alpha_2)+\frac{\nu(1-a)}{1-\nu}\right]^3.$$

This term is negative and, thus, $Q_r(\beta_r)$ is concave in stage *M*, if and only if

$$\frac{(1-\alpha_2+2\alpha_1-2a)(1-\alpha_2)}{2(\alpha_2-\alpha_1+a)(1-a)} < \frac{\nu}{1-\nu}.$$

Finally, in diffusion stage *H*, $\alpha_1 < \alpha_2 < \alpha_3 < 1$ and substitution for $\frac{\partial \alpha_1}{\partial \beta_r}, \frac{\partial \alpha_2}{\partial \beta_r}, \frac{\partial \alpha_3}{\partial \beta_r}$ yields

$$\frac{\partial^2 Q_r}{\partial \beta_r^2}\frac{z^3 Q_r}{216\nu\gamma(1-a)} = \left[(\alpha_3-\alpha_2)^2 + 2\alpha_3\frac{\nu(1-a)}{1-\nu}\right]^3 + \left[(\alpha_3-\alpha_2)^2 - 2\alpha_2\frac{\nu(1-a)}{1-\nu}\right]^3 + 8(\alpha_1-a)^3\left[(\alpha_3-\alpha_2)+\frac{\nu(1-a)}{1-\nu}\right]^3,$$

which is strictly positive due to $2\alpha_3 > 2\alpha_2$. Thus, $Q_r(\beta_r)$ is convex in diffusion stage *H*.

E Proof of Proposition 4

First, consider the case $aQ_r + Q_f > x_n$ for which $\zeta = 0$ by complementary slackness. Noting that the proposition relates to diffusion stage *FR* and taking into account that for $\alpha \in [a, \alpha_c]$ the price is p_c, the first-order conditions (19) to (21) for optimal capacities and non-reactive demand become (superscript c denotes welfare with a price cap)

$$W^c_{Q_f} = \int_a^{\alpha_c} (p_c - b_f)\, dF(\alpha) + \int_{\alpha_c}^{\alpha_2} (p_v(\alpha) - b_f)\, dF(\alpha) - \beta_f = 0, \tag{61}$$

$$W^c_{Q_r} = \int_a^{\alpha_c} (p_c - b_r)\, \alpha dF(\alpha) + \int_{\alpha_c}^{1} (p_v(\alpha) - b_r)\, \alpha dF(\alpha) - \beta_r = 0, \tag{62}$$

$$W^c_{x_n} = p_n - \int_a^{\alpha_c} p_c dF(\sigma) - \int_{\alpha_c}^{1} p_v(\alpha)\, dF(\alpha) = 0. \tag{63}$$

The procedure to determine the comparative statics is similar to that in the proof of Proposition 2. As there we denote the second-order derivatives by $W^c_{ij}; i,j = r,f,n,p$, where the two subscripts represent the variables of differentiation, Q_r, Q_f, x_n, and p_c (e.g. $W^c_{rp} := \frac{\partial W^c_{Q_r}}{\partial p_c}$). Moreover, using $\alpha_c < \alpha_1$ and $p_c = p_{v1}(\alpha_c)$ the first-order derivatives are again continuous so that the effects at the borders of the integrals cancel out when differentiating. Applying the implicit function theorem to the above equation system yields

$$\begin{pmatrix} \frac{\partial Q_r}{\partial p_c} \\ \frac{\partial Q_f}{\partial p_c} \\ \frac{\partial x_n}{\partial p_c} \end{pmatrix} = - \begin{pmatrix} W^c_{rr} & W^c_{rf} & W^c_{rn} \\ W^c_{fr} & W^c_{ff} & W^c_{fn} \\ W^c_{nr} & W^c_{nf} & W^c_{nn} \end{pmatrix}^{-1} \begin{pmatrix} W^c_{rp} \\ W^c_{fp} \\ W^c_{np} \end{pmatrix}. \tag{64}$$

Observe that the respective first integrands in equation system (61) to (63) are independent of Q_r, Q_f, x_n. Moreover, if one replaces in the respective second integrands the lower bound α_c by a, then the other terms are the same as in the three (binding) first-order conditions (19) to (21) for $\zeta = 0$. Therefore, the second-order derivatives in the Hessian matrix, denoted A^c, in (64) are the same as those in (40) and, thus, the same as given by (42) to (46) if one always replaces a by α_c in the numerator (note that the term $1-a$ in the denominator is the density and remains unchanged). For the same reason, the determinant $|A^c|$ of the Hessian in (64) is the same as the determinant $|A|$ of the Hessian in (40) and, thus, as given by (47) if one always replaces a by α_c in the numerator. Noting that $\alpha_c - \alpha_1 < 0$, it follows immediately that $|A^c| < 0$.

The inverse of the matrix A^c can be determined by $(A^c)^{-1} = \frac{1}{|A^c|}\mathrm{Adj}(A^c) = \frac{1}{|A^c|}\mathrm{Cof}(A^c)^T$ so that the expressions for $\frac{\partial Q_r}{\partial p_c}$ and $\frac{\partial Q_f}{\partial p_c}$ in (64) become (i.e. we omit $\frac{\partial x_n}{\partial p_c}$ for parsimony)

$$\begin{aligned} \begin{pmatrix} \frac{\partial Q_r}{\partial p_c} \\ \frac{\partial Q_f}{\partial p_c} \end{pmatrix} &= -\frac{(\alpha_c - a)}{(1-a)|A^c|} \begin{pmatrix} A^c_{11} & A^c_{21} & A^c_{31} \\ A^c_{12} & A^c_{22} & A^c_{32} \end{pmatrix} \begin{pmatrix} \frac{1}{2}(\alpha_c + a) \\ 1 \\ -1 \end{pmatrix} \\ &= -\frac{(\alpha_c - a)}{(1-a)|A^c|} \begin{pmatrix} \frac{1}{2}(\alpha_c + a) A^c_{11} + A^c_{21} - A^c_{31} \\ \frac{1}{2}(\alpha_c + a) A^c_{12} + A^c_{22} - A^c_{32} \end{pmatrix}, \end{aligned} \tag{65}$$

where we have used that $W^c_{rp} = \int_a^{\alpha_c} \alpha dF(\alpha) = \frac{\alpha_c^2 - a^2}{2(1-a)}$, $W^c_{fp} = -W^c_{np} = \int_a^{\alpha_c} dF(\alpha) = \frac{\alpha_c - a}{1-a}$, and the entries of the cofactor matrix $\mathrm{Cof}(A^c)$ are

$$A^c_{11} = \begin{vmatrix} W^c_{ff} & W^c_{fn} \\ W^c_{nf} & W^c_{nn} \end{vmatrix} = \frac{(\alpha_3 - \alpha_2)(\alpha_1 - \alpha_c) + \frac{\nu(1-a)}{1-\nu}(\alpha_1 - \alpha_c)}{\nu^2 \gamma^2 (1-a)^2},$$

$$A^c_{12} = A^c_{21} = -\begin{vmatrix} W^c_{fr} & W^c_{fn} \\ W^c_{nr} & W^c_{nn} \end{vmatrix} = \frac{\left(\alpha_3^2 - \alpha_2^2\right)(\alpha_1 - \alpha_c) - (\alpha_3 - \alpha_2)\left(\alpha_1^2 - \alpha_c^2\right) - \frac{\nu(1-a)}{1-\nu}\left(\alpha_1^2 - \alpha_c^2\right)}{2\nu^2\gamma^2(1-a)^2},$$

$$A_{31}^c = \begin{vmatrix} W_{fr}^c & W_{ff}^c \\ W_{nr}^c & W_{nf}^c \end{vmatrix} = \frac{\frac{1}{2}\left(\alpha_3^2 - \alpha_2^2\right)(\alpha_1 - \alpha_c)}{\nu^2\gamma^2 (1-a)^2},$$

$$\begin{aligned} A_{22}^c = \begin{vmatrix} W_{rr}^c & W_{rn}^c \\ W_{nr}^c & W_{nn}^c \end{vmatrix} &= \frac{(\alpha_3 - \alpha_2)^4 + (\alpha_1 - \alpha_c)^4 + 4\left(\alpha_1^3 - \alpha_c^3\right)(\alpha_3 - \alpha_2) + 4\left(\alpha_3^3 - \alpha_2^3\right)(\alpha_1 - \alpha_c)}{12\nu^2\gamma^2(1-a)^2} \\ &+ \frac{4\frac{\nu(1-a)}{1-\nu}\left(\alpha_3^3 - \alpha_2^3 + \alpha_1^3 - \alpha_c^3\right) - 6\left(\alpha_3^2 - \alpha_2^2\right)\left(\alpha_1^2 - \alpha_c^2\right)}{12\nu^2\gamma^2(1-a)^2}, \end{aligned}$$

$$A_{32}^c = -\begin{vmatrix} W_{rr}^c & W_{rf}^c \\ W_{nr}^c & W_{nf}^c \end{vmatrix} = \frac{\frac{1}{3}\left(\alpha_3^3 - \alpha_2^3\right)(\alpha_1 - \alpha_c) - \frac{1}{4}\left(\alpha_3^2 - \alpha_2^2\right)\left(\alpha_1^2 - \alpha_c^2\right) + \frac{1}{12}(\alpha_1 - \alpha_c)^4}{\nu^2\gamma^2(1-a)^2}.$$

Accordingly,

$$\frac{1}{2}(\alpha_c + a)A_{11}^c + A_{21}^c - A_{31}^c = -\frac{\frac{1}{2}(\alpha_1 - \alpha_c)(\alpha_1 - a)\left(\alpha_3 - \alpha_2 + \frac{\nu(1-a)}{1-\nu}\right)}{\nu^2\gamma^2(1-a)^2} < 0,$$

so that upon substitution into (65) we obtain $\frac{\partial Q_r^c}{\partial p_c} < 0$ for all diffusion stages. Given that capacities without a price cap are efficient, it follows immediately that $Q_r^c > Q_r^*$. Moreover,

$$\begin{aligned} \frac{1}{2}(\alpha_c + a)A_{12}^c + A_{22}^c - A_{32}^c &= \frac{(\alpha_3 - \alpha_2)\left[(\alpha_3 - \alpha_2)^3 + (\alpha_1 - \alpha_c)^3 + 3\left(\alpha_1^2 - \alpha_c^2\right)(\alpha_1 - a)\right]}{12\nu^2\gamma^2(1-a)^2} \\ &- \frac{3\left(\alpha_3^2 - \alpha_2^2\right)(\alpha_1 - \alpha_c)(\alpha_1 - a)}{12\nu^2\gamma^2(1-a)^2} \\ &+ \frac{\nu(1-a)}{1-\nu}\frac{4\left(\alpha_3^3 - \alpha_2^3\right) + (\alpha_1 - \alpha_c)^3 + 3\left(\alpha_1^2 - \alpha_c^2\right)(\alpha_1 - a)}{12\nu^2\gamma^2(1-a)^2}. \end{aligned} \tag{66}$$

In diffusion stages V and L, for which $\alpha_3 = \alpha_2 = 1$, the terms in the first two lines cancel and $\frac{\partial Q_f^c}{\partial p_c} > 0$. By contrast, in diffusion stages M and L, the sign of $\frac{\partial Q_f^c}{\partial p_c}$ is inconclusive because for $\alpha_3 > \alpha_2$ there is a negative term in the second line of (66). However, because $\lim_{\nu \to 1} \frac{\nu}{1-\nu} = \infty$ the positive term in the third line unambiguously dominates and $\frac{\partial Q_f^c}{\partial p_c} > 0$ for ν sufficiently large.

Next, consider the case $aQ_r + Q_f = x_n$. The procedure is again similar to the corresponding case in the proof of Proposition 2. The only effect of a price cap on the first-order conditions (19) to (21) for Q_r, Q_f, x_n is that the integral terms have to be split into two parts. The first applies to $\alpha \in [a, \alpha_c]$ with a fixed price p_c, while the second applies to $\alpha \in (\alpha_c, 1]$ and remains unchanged otherwise. Doing so, solving (21) for ζ, substituting this into (19) and (20) and using $x_n = aQ_r + Q_f$, the first-order conditions are the same as (48) and (49), if one replaces $\int_a^1 p_v(\alpha)(\alpha - a)dF(\alpha)$ in the latter condition by $\int_a^{\alpha_c} p_c(\alpha - a)dF(\alpha) + \int_{\alpha_c}^1 p_v(\alpha)(\alpha - a)dF(\alpha)$.

Upon differentiation, $W_{fp}^c = 0$, which reflects that $p_c > b_f$ so that, whenever the price cap binds, it is cheaper for fossil firms to produce the electricity needed for its non-reactive consumers themselves, rather than buying it at the price p_c. Moreover,

$$W_{rp}^c = \int_a^{\alpha_c} (\alpha - a)\, dF(\alpha) = \frac{(\alpha_c - a)^2}{2(1-a)}$$

so that applying the implicit function theorem to the system of equations $W_{Q_r}^c = W_{Q_f}^c = 0$ yields

$$\begin{pmatrix} \frac{\partial Q_r}{\partial p_c} \\ \frac{\partial Q_f}{\partial p_c} \end{pmatrix} = -\begin{pmatrix} W^c_{rr} & W^c_{rf} \\ W^c_{fr} & W^c_{ff} \end{pmatrix}^{-1} \begin{pmatrix} W^c_{rp} \\ 0 \end{pmatrix} \tag{67}$$

$$= -\frac{(\alpha_c - a)^2}{2(1-a)\,|B_c|} \begin{pmatrix} W^c_{ff} \\ -W^c_{rf} \end{pmatrix}, \tag{68}$$

The second-order derivatives $W^c_{ff} < 0$ and $W^c_{rf} = W^c_{fr}$ are the same as the corresponding expressions (51) to (52) (because α_c does not appear at the border of the integral in these expressions), while[16]

$$\begin{aligned} W^c_{rr} &= -\int_{\alpha_c}^{\alpha_1} \frac{(\alpha - a)^2}{\nu\gamma} dF(\alpha) - \int_{\alpha_2}^{\alpha_3} \frac{(\alpha - a)^2}{\nu\gamma} dF(\alpha) - \frac{a^2}{(1-\nu)\gamma} \\ &= -\frac{(\alpha_1 - a)^3 - (\alpha_c - a)^3}{3(1-a)\nu\gamma} - \frac{(\alpha_3 - a)^3 - (\alpha_2 - a)^3}{3(1-a)\nu\gamma} - \frac{a^2}{(1-\nu)\gamma} < 0. \end{aligned} \tag{69}$$

Upon substitution, the determinant of the Hessian is

$$|B^c| = \frac{\frac{1}{4}(\alpha_3 - \alpha_2)^4 + \left[(\alpha_1 - a)^3 - (\alpha_c - a)^3\right]\left(\alpha_3 - \alpha_2 + \frac{\nu(1-a)}{1-\nu}\right) + \frac{\nu(1-a)}{1-\nu}\left(\alpha_3^3 - \alpha_2^3\right)}{3\nu^2\gamma^2(1-a)^2},$$

which is clearly positive due to $\alpha_1 - \alpha_c > 0$. Upon substitution into (68), it follows immediately that $\frac{\partial Q_r}{\partial p_c} > 0$ and $\frac{\partial Q_f}{\partial p_c} < 0$ if and only if

$$W^c_{rf} = \frac{\alpha_3^2 - \alpha_2^2}{2(1-a)\nu\gamma} - a\frac{\alpha_3 - \alpha_2}{(1-a)\nu\gamma} - \frac{a}{(1-\nu)\gamma} < 0.$$

This is the case in diffusion stage V and L for which $\alpha_3 = \alpha_2 = 1$. By contrast, in diffusion stage M and H the sign is inconclusive due to the positive first term. However, $\lim_{\nu \to 1} \frac{a}{1-\nu} = \infty$ implies that the third, negative term unambiguously dominates and $\frac{\partial Q^c_f}{\partial p_c} < 0$ for ν sufficiently large.

[16] The difference results because the term that corresponds to $\alpha_c - a$ in (69) is $a - a = 0$ in (53).