CESifo WORKING PAPERS

8673
2020

Original Version: November 2020
This Version: March 2021

Efficiency in Wholesale Electricity Markets: On the Role of Externalities and Subsidies

Sylwia Bialek, Burcin Unel

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8673*

Efficiency in Wholesale Electricity Markets: On the Role of Externalities and Subsidies

Abstract

This paper studies the effects of capacity market reforms that the U.S. grid operators have undertaken in response to state-level subsidies paid to emission-free electricity generation. We first derive an analytical model of energy-and-capacity markets that allows us to predict the price and resource mix effects of subsidies, as well as to understand their welfare implications. We confirm that while the subsidies, even when combined with energy consumption taxes, generally cannot achieve first-best outcomes. We also show there exists a range of subsidy rates that are welfare-enhancing when greenhouse gas externalities are taken into account. Finally, we focus on the evaluation of capacity market reforms similar to those in PJM, ISO-NE, and NYISO, finding that such reforms are likely to decrease welfare.

JEL-Codes: Q280, Q420, Q580, Q410.

Keywords: capacity markets, renewables, subsidies, welfare.

*Sylwia Bialek**
Institute for Policy Integrity
New York University / USA
sylwia.bialek@nyu.edu

Burcin Unel
Institute for Policy Integrity
New York University / USA
burcin.unel@nyu.edu

*corresponding author

March 2021

1 Introduction

In the U.S., the increasing prevalence and magnitude of clean energy subsidies provided by the states in the form of Renewable Energy Credits and Zero-Emission Credits[1] raised worries about their effect on capacity markets—markets which complement energy markets in some regions to ensure adequate electricity supply. In particular, a concern arose that such generation subsidies could harm market efficiency by suppressing capacity-market prices below their competitive levels. The concern triggered controversial capacity market reforms that intended to mitigate the impacts of state subsidies within three U.S. electricity trading regions: PJM, New York-ISO (NYISO) and ISO-New England (ISO-NE).[2] As the regions that experienced reforms constitute a significant part of the U.S. electricity system,[3] grid decarbonization efforts could be largely obstructed if the reforms reach their stated goal of "mitigating" the effects of subsidies for clean generation.This possibility led to ongoing litigation, and even started discussion in some state about leaving the capacity markets all together.

Given the prevalence of generation subsidies and the policy concerns around them, it is crucial to understand the consequences of subsidizing pollution-free electricity generation. Such consequences will generally depend not only on the design of the subsidy scheme [Fell and Linn (2013), Abrell et al. (2019)] but also on the organization of the electricity systems. The mechanisms through which subsidies affect final outcomes are different in settings with vertically integrated utilities than in systems with wholesale electricity markets. And within the electricity markets, it is conceivable that the impact channels depend on the market design, for instance on whether the market is operated as energy-only or energy-plus-capacity market. And while cost-effectiveness of subsidies has been studied intensively in the context of production of goods [Jung et al. (1996), Acemoglu et al. (2012)], the interaction of generation subsidies with the design of wholesale electricity markets is still little understood.

Our main goal in this paper is to assess whether, and under what conditions, generation subsidies can lead to an increase in the economic efficiency of wholesale electricity markets. We consider only subsidies for generators that produce electricity pollution-free because of

[1] As of 2020, twenty-nine states have a Renewable Portfolio Standard. New York and Illinois pay some of their nuclear generators Zero-Emission Credits, while five further states are considering similar payments. Moreover, the federal Renewable Electricity Production Tax Credit provides another per-MWh subsidy to renewable generators.

[2] For the reform descriptions, see the discussion in Section 5 as well as the Orders 162 FERC ¶ 61,205 (2018), 169 FERC ¶ 61,239 (2019) and 172 FERC ¶ 61,058 (2020) issued by Federal Energy Regulatory Commission.

[3] In total, the three trading regions serve a third of all American electricity customers and host one fourth of the total 1,100 GW of generation capacity installed.

relevance of such subsidies for policy discussions and the problem of uninternalized externalities in electricity markets.[4] Given the recent reforms in PJM, NYISO, and ISO-NE, we are particularly interested in understanding if and how generation subsidies could affect capacity market prices. To answer these questions, we derive an analytical, partial-equilibrium model of wholesale energy and capacity markets building on the seminal study by Joskow and Tirole (2007). We determine the effect that generation subsidies have on the equilibrium prices and generation mix. We then compare the welfare outcomes under generation subsidies to two benchmarks: the status quo, where there is no tax on greenhouse gas emissions; and the "first-best" case, where there is a Pigouvian tax on externalities.

Despite the policy relevance, economic research on the impact of generation subsidies on the efficiency of wholesale electricity markets has been scarce. Most studies on the issue consider only private generation costs [Briggs and Kleit (2013), Brown (2018a), Brown (2018b), Blumsack et al. (2018), Llobet and Padilla (2018)], even though assuming away externalities leaves no room for the subsidies to be welfare enhancing. Other studies sidestep the question of efficiency, concentrating on the individual effects of the subsidies [Bento et al. (2018), Haan and Simmler (2018), Abrell, Kosch, and Rausch (2019)]. Yet others look at the costs and benefits of the subsidy policies by simulating their results in a complex electricity market model, often using energy-only models [Palmer et al. (2011), Fell and Linn (2013), Reguant (2019), Abrell, Rausch, and Streitberger (2019)]. While such studies give us a good grasp on the net effects of a particular policy, they typically cannot provide information about the underlying mechanisms and, therefore, do not easily lend themselves to generalization of the results.[5] Finally, some analyses are conducted from the perspective of a social planner, abstracting away from wholesale electricity markets [Antoniou and Strausz (2017), Eichner and Runkel (2014)] even though the wholesale market design can render some allocations infeasible.

Abstracting away from the design of wholesale electricity markets, such as the existence of capacity markets, limits the researchers' ability to account for the interaction between the energy and capacity markets, and thus hinder their ability to evaluate long run equilibrium outcomes in these markets. In regions like PJM, capacity markets are an important source of revenue for generators with total annual capacity payments corresponding to around one third of energy payments.[6] Further, not accounting for the interaction between the markets and analyzing them separately might lead to misguided welfare conclusions.

[4]Subsidies that are not a response to a market failure will by definition be distortive.

[5]The exception is Fell and Linn (2013) who use a simple investment model that cleanly delineates the intuition behind intermittency and cost-effectiveness of different subsidy policies. However, the model does not account for many basic features of energy markets and abstracts from the existence of capacity markets.

[6]In the 2018/2019 capacity year, the total capacity payments in PJM were $11.0 billion (IMM, 2019).

Our first contribution in this paper is to fill this gap in understanding the interaction between generation subsidies and wholesale energy and capacity markets. We derive changes in long-run equilibrium energy mix and energy and capacity prices spurred by subsidies. The novelty of our results is twofold here. First, to the best of our knowledge, our paper is the first to show analytically the relationship between generation subsidies and equilibrium capacity market prices. Additionally, by explicitly accounting for fluctuations in demand for electricity, we are able to show the heterogeneity within the energy price effects of subsidies. Papers that assume one electricity price (e.g. Fischer (2010)) cannot account for the fact that the peak prices are affected differently by subsidies than low demand prices. And such distinction is important when there is heterogeneity in pollution intensity of the fossil-fueled generators or when the price sensitivity of demand varies between demand levels.

Our second contribution is providing an analytical framework for analyzing market effects of various reforms. The framework allows researchers to incorporate the desired heterogeneity of generation – including heterogeneity in emission intensities – and demand variability while allowing for transparent, closed-form solutions for energy and capacity prices, and capacity levels. The models also accounts for the inter-dependency between energy and capacity markets. This is especially important given that regulators are currently rethinking the design of electricity markets for a decarbonized future.

As a third contribution, we provide a welfare assessment of subsidies in settings with wholesale electricity markets and externalities. We outline the conditions under which subsidies can be welfare-improving. We demonstrate that while subsidies alone cannot produce the first-best outcomes that could be achieved by a Pigouvian tax, there always exists a range of welfare-enhancing subsidy rates when pollution damages are ignored by the market participants and subsidies are financed from a general budget. For subsidies financed through additional charges on electricity consumption, there exist demand and supply technology configurations where any subsidy rate reduces economic efficiency. And while the general welfare results for subsidies are already known from the literature on subsidies (e.g. Holland et al. (2009)), the refinement in our paper comes from accounting for properties of electricity markets. In particular, varying electricity demand and heterogeneity in pollution intensity of dirty generation create additional welfare effects.

Finally, we analyze the welfare implications of the recent capacity market reforms in the U.S.. We show that, since the capacity market price suppression through generation subsidies can happen only under limited circumstances, policies that indiscriminately mitigate any subsidy, without taking its welfare effect into account, will harm the economic efficiency of wholesale electricity markets.

Our results are highly topical given the prevalence of generation subsidies and the increasingly ambitious clean energy targets that will lead to an even higher share of subsidized resources. We offer new insights about the optimal generation subsidies for policy-makers pursuing increased generation from non-polluting resources. At the same time, we provide guidance for energy market regulators and operators on how these subsidies should be treated in the market.

The rest of the paper is divided as follows. Section 2 sets the modeling framework. In Sections 3 and 4, we compare welfare associated with market outcomes under three different regimes: no environmental policies ("status quo"), a Pigouvian tax, and generation subsidies. Section 3 focuses on energy-only markets, while Section 4 investigates energy-plus-capacity designs. Section 5 investigates the policy implications of our findings in the context of capacity market reforms in PJM, NYISO and ISO-New England triggered by generation subsidies for non-polluting resources. Section 6 concludes.

2 Electricity markets model - Preliminaries

To understand the effects of generation subsidies, we set up a model representing the long-term equilibrium in electricity markets under two market designs, an energy-only and an energy-plus-capacity market design. For the modeling, we use the framework of Joskow and Tirole (2007), but modify it in a number of dimensions.

First, we incorporate externalities and subsidies into the model. Second, to enhance tractability, we restrict ourselves to settings that allow closed-form solutions. Therefore, we assume the number of existing generation technologies and the number of possible demand states to be finite but arbitrary. Such a choice allows us to produce analytical insights on environmental policies under different resource configurations, and to capture the interactions between resource heterogeneity and environmental policies. Third, we explicitly model price formation in the wholesale energy and capacity markets to tease out the effect of generation subsidies on the markets, and thereby learn what price signals the different resources receive.

We assume there are M distinct, deterministic technologies available for electricity generation, indexed by j and contained in set $\mathcal{M}$.[7] The technology-specific useful life defines the maximum number of periods under which a given technology is physically able to operate. The technologies are also characterized by the investment cost of building 1MW of capacity, by pollution footprint expressed as damages from generation, e_j, and by a private marginal

[7] The resources' temporal availability affects their environmental and market impacts [Fell and Linn (2013), Abrell et al. (2019)], therefore it also changes the optimal subsidies and welfare. However, it does not influence the logic behind the mechanisms that we uncover. Therefore, while generation intermittency can be incorporated in our model, we abstract from it for the sake of tractability.

cost of generation, c_j^p.[8] We will refer to the marginal costs that a generator of type j faces as c_j. When there is no instrument that internalizes the externalities, c_j equals the private marginal cost, c_j^p, and when there is a Pigouvian tax it equals the social marginal cost of generation, $c_j = c_j^s = c_j^p + e_j$. We index the available technologies such that they are ordered according to their relevant marginal costs, i.e. $\forall j \in \mathcal{M}$, $c_j \leq c_{j+1}$. There are no transmission constraints or line losses in our model, so the location of generators is irrelevant.

There are N different states of demand for electricity, indexed by i contained in set $\mathcal{N}$. The states of demand are indexed in accordance with increasing demand, such that

$$\forall i, k \in \mathcal{N}, \quad i > k \implies D_i(p) = D_k(p) + \epsilon_{ik}(p),$$

where $\epsilon_{ik}(p)$ is a positive function with sufficiently high values such that the problem of "shifting peak" as described by Steiner (1957) does not emerge. In regular time intervals (e.g. 5 minutes) a new demand state is randomly drawn and the probability of a state i occurring in a given time interval equals $f_i \in (0, 1)$.[9]

There are two types of consumers — consumers with real-time meters who face state dependent prices, p_i, and consumers with traditional meters who always face a constant, pre-defined price p for a unit of electricity.[10] The state-specific demands are denoted respectively by $\hat{D}_i(p_i)$ and $D_i(p)$, and the load profiles of all consumers are identical up to a scale factor.

The market operator runs a separate standard uniform price auction for every time period to balance the electricity supply and demand. Utilities bid in the demand curve corresponding to the demand state prevailing in a given time on behalf of their consumers. We assume that utilities pass through the wholesale prices to consumers on real-time meters. For other consumers, utilities impose a constant, volumetric energy charge and a fixed charge. Generators bid into the auction their whole generation capability at the marginal cost of generation they face. As the supply curve is a step function, the clearing price can be above the price bid by marginal generator – situation referred to as "scarcity pricing."

It is straightforward to see that a generator that clears the energy auction in demand state i also clears with higher demand levels, i.e. in states with indices higher than i. We refer to a generator as an "i^{th}-merit" generator if it clears the energy auction in state i

[8] The private marginal cost should be understood as representing fuel costs, variable labor costs, etc. if the plant were to run on its full capacity for its full lifetime. The modeling framework abstracts from the start-up costs, minimum load considerations etc.

[9] For simplicity, we ignore ramping costs and constraints.

[10] An optimal contract with consumers on traditional meters includes probabilities of the consumers being rationed in individual states. However, when the marginal generation costs during peak demand are relatively low compared to the marginal surplus from peak electricity consumption for consumers on traditional meters, the optimal rationing probability is zero. To simplify the exposition, we thus assume that the above condition is met and abstract away from consumer rationing. For discussion of interruptibility, see Joskow and Tirole (2007).

but does not clear the market in state $i-1$. The capacity utilization rate of an i^{th}-merit resource is thus given by the sum of frequencies of the states when resource clears the market, $\sum_{k=i}^{N} f_k$. To ensure the uniqueness of equilibria, we assume that no two technologies have the same levelized costs of energy at the possible capacity utilization rates.[11]

In the set of technologies clearing the auction in a given state demand state, we call the technology with the highest marginal cost the marginal technology for that state. The merit order is defined as mapping h assigning a marginal technology to demand state, $h : \mathcal{N} \mapsto \mathcal{M}$.

The market operator may also want run a capacity market in addition to energy markets. The supply curve in capacity markets is formed through generator bids representing the minimum price they are willing to accept for keeping their capacity ready to produce electricity in a given time period (e.g. one year).[12] We model this market as a uniform price auction where the amount of capacity procured corresponds to the predicted maximum amount of demand.[13]

3 Energy-Only Markets

In this Section, we analyze the effects of subsidies on energy-only markets, like ERCOT in Texas. In the next Section, we extend the analysis to understand how capacity markets interact with generation subsidies.

3.1 General findings on energy supply and equilibrium energy prices

Below we present a two-step procedure for establishing market equilibrium with an energy-only market design. In the first step, we identify the set of technologies that are economic and therefore belong into the equilibrium generation mix. In the second step, we derive the equilibrium prices and energy consumption in individual states.

Let $\mathcal{P}_{\mathcal{N}}$ denote the m-element subset of generation technologies that, given the distribution of states, f_i, clear the energy auction in at least one demand state. In what follows, we

[11] Levelized cost of energy represents the average revenue per unit of electricity generated that would be required to recover the costs investment and generation given the share of time that the generator clears the market. The assumption of unique levelized costs eliminates the possibility of two technologies looking the same from the market perspective.

[12] The clearing generators are allowed to also participate in the energy market. In other words, the market is not of "reserve" type like in Germany and Belgium.

[13] In the U.S., the Midcontinent Independent System Operator, among others, uses this type of capacity market. Other U.S. capacity auctions use capacity demand curves, which complicate the exposition but do not change the underlying mechanisms studied here.

will refer to those technologies as "economic."[14] We make the following observation about the relationship between the demand states and economic types of generators:

Lemma 1. *In a competitive electricity market with technologies that have constant returns to scale and unique levelized costs of entry, the mapping from the demand states to the set of economic technologies that defines the merit of the economic generators, $g : \mathcal{N} \mapsto \mathcal{P}_{\mathcal{N}}$, is surjective. Consequently, the set $\mathcal{P}_{\mathcal{N}}$ contains at most $m = N$ elements.*

Proof - see appendix A.1. □

Under these assumptions, the equilibrium resource mix and energy prices in individual states remain constant over time, despite the individual generators exiting the market as a result of their limited useful life. As generators retire at the end of their useful life, they get swiftly replaced by a generator of the same technology, leading to a constant equilibrium capacity mix and prices. As the useful life solely redefines the cost structure allowing the recovery of investment costs to be spread over multiple years, we can investigate optimality in an individual period and focus on annualized investment costs, I_j.

Lemma 2. *The marginal type of generator j in state i can be defined s.t.*

$$I_j + c_j \sum_{k=i}^{N} f_k \leq I_g + c_g \sum_{k=i}^{N} f_k \quad \forall g \in \mathcal{M}. \tag{1}$$

Proof - see appendix A.2. □

Condition (1) establishes which technologies are economic given the distribution of the demand states and the pollution internalization mechanism (or lack thereof), revealing the set of economic technologies, $\mathcal{P}_{\mathcal{N}}$, concluding the first step of the procedure.

The second step of our procedure establishes the equilibrium prices and consumption levels given the economic technology mix. We re-index all the economic technologies belonging to $\mathcal{P}_{\mathcal{N}}$ such that their indices correspond to the states in which they are marginal. Note that with re-indexing, it still true such that generators with lower marginal costs are lower in merit order, such that $\forall j \in \mathcal{P}_{\mathcal{N}}$, $c_j \leq c_{j+1}$. However, this re-indexing allows us to skip technologies that are never economic given the demand states. Note also that technologies that are marginal in multiple subsequent states are indexed at least twice, i.e. i^{th}-merit and $i+1^{th}$-merit resources might represent the same technology.

Assume that markets and utilities account for externalities only to the extent that they are internalized through emission pricing, i.e. $c_j = c_j^p + tax$. Given this assumption and

[14] Note that with a Pigouvian tax in place, the definition a generator being economic also accounts for externalities.

$\mathcal{P}_{\mathcal{N}}$, prices and quantities that emerge in the market are the same as prices and quantities obtained using the following Lagrangian:

$$\begin{aligned}\max_{p,\,\{p_j\}_{j\in\mathcal{N}}} \quad & \sum_{i=1}^{N} f_i[S_i(p)) + \hat{S}_i(p_i) - \sum_{j=1}^{i} c_j K_j] - \sum_{i=1}^{N} I_i K_i \\ & - \sum_{i=1}^{N} \lambda_i (K_i - D_i(p) - \hat{D}_i(p_i) + D_{i-1}(p) + \hat{D}_{i-1}(p_{i-1})),\end{aligned} \tag{2}$$

with $\hat{D}_0 = 0$ and $D_0 = 0$, K_j denoting the capacity, i.e. number of MWs built of technology j, $S_i(p)$ and $\hat{S}_i(p_i)$ representing the gross surplus of consumers on traditional meters and real-time meters, respectively, and λ_i reflecting the shadow price of capital constraint in demand state i. This maximization leads to the following equilibrium prices, which are a generalization of Joskow and Tirole (2007) for a discrete number of states:

$$p_i^E = \frac{c_i \sum_{k=i}^{N} f_k - c_{i+1} \sum_{k=i+1}^{N} f_k + I_i - I_{i+1}}{f_i} \quad i = 1, 2...N-1 \tag{3}$$

$$p_N^E = c_N + \frac{I_N}{f_N}, \tag{4}$$

$$p^E = \frac{\sum_{i=1}^{N} f_i p_i^E D_i'}{\sum_{i=1}^{N} f_i D_i'}. \tag{5}$$

While competitive generators bid their marginal energy costs, equilibrium revenue they get should allow them to break even. Given that the generators of type N sell energy only in the peak period, the per MWh energy peak price, p_N^E, needs to cover their marginal costs of energy generation, c_N, and their annualized cost of investment adjusted by the frequency of peak demand occurrence, $\frac{I_N}{f_N}$. Prices above that level would induce new entry of generation, until, through increased supply of energy and downward-sloping demand, the price falls to p_N^E. The mid-merit generator makes energy sales in intermediate-peak and peak periods. Given p_N^E, the equilibrium p_{N-1}^E needs thus to meet the following condition: $I_{N-1} + (f_{N-1} + f_N)c_{N-1} = f_{N-1}p_{N-1}^E + f_N p_N^E$. Similar logic of "backward induction" can be used to obtain p_{N-2}^E, p_{N-3}^E, etc.

The deviations of the equilibrium prices from marginal costs represent scarcity pricing needed to equate demand and supply. The magnitude of the deviations decreases as the number of demand states. Note also that if a technology is marginal in more than one state, the prices in the states with lower indices equal its marginal generation cost.

The amount of capacity of each resource type is determined by the differences in demand levels between individual states, i.e. $K_i^E = D_i(p^E) + \hat{D}_i(p_i^E) - D_{i-1}(p^E) - \hat{D}_{i-1}(p_{i-1}^E)$. The results assume full divisibility of the generation units. Lumpiness of capacity investments complicates the formula, and detracts from efficiency as shown in Antoniou and Strausz (2017), but does not change the main insights from our model.[15]

The incentive-optimal price for consumers on traditional meters, p^E, does not guarantee cost recovery to the utility for providing electricity to those consumers. Therefore, the utility recovers (or returns) the missing (or excess) energy revenue from (to) consumers on traditional meters through the use of fixed charges.

If a utility is concerned about the external damages, it might want to deviate from pricing formula 5 even when the Pigouvian taxes are absent and implement its pricing for consumers on traditional meters using socially optimal price. Given that the utility can use a multiplicity of potential pricing designs for consumers on traditional meters, based on how it accounts for externalities in its objective function, we relegate the results relevant for those consumers to the appendix B and in the reminder of the paper we focus solely on consumers on real-time meters.

3.2 Comparing outcomes under the status quo and Pigouvian tax

Under the status quo, there are no instruments that fully internalize external damages from emissions. Consequently, the wholesale prices obtained in formulas (3)-(4) (which also correspond to retail prices for consumers on real-time meters) reflect private and not social generation costs, leading to distortions in market outcomes. We call those distortions "intensive" when the quantities produced by individual generator types are inefficient but the merit order, h, is the same as under the "first-best" Pigouvian tax,[16] such that the existence of externalities does not change the types of technologies that are "economic" and clear the market. A sufficient condition for the distortions having intensive character reads:

$$\forall i, j \in \mathcal{M}, t \in \mathcal{N}, \;\; I_i + c_i \sum_{k=t}^{N} f_k \leq I_j + c_j \sum_{k=t}^{N} f_k$$
$$\implies I_i + (c_i + e_i) \sum_{k=t}^{N} f_k \leq I_j + (c_j + e_j) \sum_{k=t}^{N} f_k \, .$$

[15] If there is lumpiness in investment manifesting itself in minimal size of peaker capacity extension of y, the equilibrium capacity amount might be up to y units lower than than the equilibrium capacity with divisible investment. The deviation of the associated "lumpy" peak price from p_N^E is given by: $p_N^{cl} \leq p_N^E + y[\hat{D}'_N + D'_N f_3 (D'_N - \sum_{i=1}^{N-1} D'_i)^{-1}]$. The smaller the minimum size of investment is in comparison to the demand, the lower will be the relative impact of lumpiness on the outcomes.

[16] When some consumers are on traditional meters and thus do not respond to real time prices, first best cannot be reached. Therefore, even the Pigouvian tax represents a second-best world.

When externalities change the merit order, we call the resulting distortions "extensive." If the distortions are extensive, resources that clear the market under the status quo are different than those that would clear under first-best outcomes. In the main body of the article, we focus on intensive distortions, which have closed form solutions, but we reproduce some of the results for extensive distortions in the appendix. We unify the analysis of the two types of distortions by extending a general model presented in Joskow and Tirole (2007) to account for emission damages and present it in Appendix C as a benchmark for our results.

We index the prices under the status quo approach with "SQ." Comparing the status quo prices obtained through formulas (3)-(4) with the socially optimal prices that account for social marginal costs (indexed with "$*$"), we arrive at Proposition 1.

Proposition 1. *When externalities do not change the merit order, the differences between the status quo and the socially optimal wholesale prices are as follows:*

$$p_i^{\Delta} = p_i^{SQ} - p_i^* = \frac{e_{i+1} \sum_{k=i+1}^{N} f_k - e_i \sum_{k=i}^{N} f_k}{f_i}, \quad i = 1, 2...N-1 \tag{6}$$

$$p_N^{\Delta} = p_N^{SQ} - p_N^* = -e_N \leq 0. \tag{7}$$

The status quo prices are (weakly) too high when pollution-free resources are marginal and suppressed when dirtiest resources are on the margin. For low polluting resources, the sign of price distortion depends on the pollution intensity of the technology that follows them in the merit order.

The distortions in prices translate into distortions in generation capacities, which for an i^{th}-merit resource can be written as:

$$K_i^{\Delta} = K_i^{SQ} - K_i^* = p_i^{\Delta} \hat{D}_i' - p_{i-1}^{\Delta} \hat{D}_{i-1}' \tag{8}$$

Inspecting Equations (6)-(8) it is easy to see that the distortions skew the generation mix towards polluting resources: the pollution-free resource is (weakly) underbuilt under the status quo[17] and the most polluting types of generation get overbuilt. Predicting the direction of capacity distortion for generators that have low but positive pollution requires knowledge of price sensitivity of demand and the relative magnitude of environmental externalities of the individual technologies. The realized capacity of those polluting resources that are sufficiently less polluting than the generators above or below them in the merit order is suboptimally low.

[17] The amount of capacity for a pollution-free generation type will be correct if the adjacent merit order generators are also pollution free.

3.3 Equilibrium effects of subsidies

Policymakers attempt to correct distorted market outcomes. In recent years, their attempts have concentrated on generation subsidies for non-polluting resources. We incorporate such generation subsidies into the model by replacing generator j's marginal cost,c_j, with $c_j - s_j$ whenever it obtains a per MWh subsidy s_j. We will call a subsidy "non-disruptive" if it does not change the merit order.[18]

Theorem 1. *Assume that generation technologies exhibit constant returns to scale and that investments are perfectly divisible. Granting a non-disruptive generation subsidy, s, financed from the general budget to i^{th}-merit generators decreases p_i^E by $\Delta_i(s) = \frac{s}{f_i}\sum_{k=i}^{N} f_k$ and rises p_{i-1}^E by $\Delta_{i-1}(s) = \Delta_i(s)\frac{f_i}{f_{i-1}}$, while leaving the clearing prices in other states unaffected.*

Proof - see appendix A.3. □

Note that when the subsidy is granted to the 1^{st}-merit generators, the only price affected is p_1^E. If two or more pollution-free resource types follow each other in the merit order, i.e. if both i^{th} and $(i-1)^{th}$-merit resources are pollution-free, the price adjustment for the $i-1$ state is a compounded effect of the price effect in state i and the subsidy, leading to a decrease in p_{i-1}^E equal to s.

Corollary 1. *Granting a non-disruptive generation subsidy financed from the general budget to i^{th}-merit generators increases equilibrium capacity of i^{th}-merit resource, scales down the capacity of $(i-1)^{th}$-merit and $(i+1)^{th}$-merit resources and has no impact on the capacity of other resources.*

The adjustments in capacity follow from the price adjustments described in Theorem 1. As p_{i-1}^E decreases while p_{i-2}^E remains constant, the equilibrium amount of $i-1^{th}$-merit generation drops. On the other hand, as both p_i^E and $\sum_{k=0}^{i-1} K_k$ decline, the equilibrium capacity of the subsidized resource rises. The contraction in capacity of $i+1^{th}$-merit resource derives from increased aggregate capacity in lower demand states, $\sum_{k=0}^{i} K_k$ combined with unchanged price p_{i+1}^E.[19] □

3.3.1 Comparing effects of subsidies to first-best outcomes

To think through the effects of subsidies, assume that there are three types of resources: wind, coal and gas power plants – a resource mix that is relevant for ongoing policy discus-

[18] "Disruptive" subsidies, which change the merit order, are discussed in the appendix D in the context of extensive externalities.

[19] Lumpiness of investment could lead to subsidies affecting resources that are further away in the merit order. However, as we argued previously, we expect the effects of lumpiness to be of third-order importance if the minimum investment size is relatively small compared to the demand. Therefore, we ignore it in our analysis.

sions – and that the resources are characterized by a following cost ordering: $c_w < c_c < c_g$, $e_w = 0$, and $e_c > e_g$.[20] The price distortions under the status quo are:

$$p_1^\Delta = \frac{f_2+f_3}{f_1} e_c > 0,\ p_2^\Delta = \frac{f_3(e_g - e_c) - f_2 e_c}{f_2} < 0 \text{ and } p_3^\Delta = -e_g < 0,$$

leading to unerbuilding of wind and overbuilding of coal. There will be excess gas capacity if $e_c < e_g \frac{D_3'/D_2' f_2 + f_3}{f_2+f_3}$.

Giving a subsidy $s = e_c(f_2 + f_3)$ to wind would implement the optimal price p_1^* but leave the prices p_2^E and p_3^E unaffected, preserving the distortions in capacity of the polluting resources. Alternatively, policymakers could implement a subsidy that brings the capacity of coal or gas capacity to an optimal level or use a subsidy that maximizes the welfare taking into account uniternalized pollution. Nevertheless, based on Theorem 1, it is clear that no subsidy rate can remove all the distortions.[21]

Policymakers could refine the subsidy policy by combining it with an electricity consumption tax. With two instruments, it is possible to target two outcomes, e.g. induce optimal p_1^E and p_3^E using a combination of tax $t = e_g$ and subsidy $s = e_c(f_2 + f_3)$. Nevertheless, the combined instruments cannot restore the optimality since $p_2^E \neq p_2^*$ (unless by coincidence).

Proposition 2. *If there are more than two demand states, and if there are two or more types of polluting generators that are economic, a homogeneous subsidy for pollution-free energy generation, even when combined with an electricity consumption charge, cannot produce first-best outcomes.*

With N demand states, up to N prices are distorted;[22] however, a subsidy affects only prices of pollution-free resources and prices of the resources directly below them in the merit order. Besides, a subsidy can not perfectly correct all the prices it affects, unless by accident – the distortions differ between the demand states and a homogeneous subsidy cannot target that distortion heterogeneity. Enhancing the subsidy with a consumption charge adds only one degree of freedom for the policymaker. □

3.3.2 Comparing effects of subsidies to the status quo

A subsidy for non-polluting generation, even when combined with a consumption tax, cannot achieve the first-best outcome. However, it can still be a socially desirable tool.

[20]Note that for illustration purposes, we assumed that total social marginal costs of coal are lower than those of gas, $c_c + e_c < c_g + e_g$. However, this does not need to hold true and will mostly depend on the relative prices of coal and gas.

[21]For instance, in the example used, a subsidy is capable of changing solely the off-peak price, even though the prices in all three states are distorted.

[22]As visible in formula 6, all prices are distorted unless two or more non-polluting resources follow one another in the merit order, in which case some of the prices will be correct.

Theorem 2. *When pollution is not internalized in the market but the merit order is correct, there exists a generation subsidy for non-polluting resources financed from the general budget that weakly increases efficiency of the market compared to the status quo.*

Proof - see appendix A.4. □

The intuition is that for a sufficiently small subsidy, all of the affected prices move closer to their optimal value. By the same token, there exists a threshold above which a subsidy brings all the affected prices further away from the social optimum. For subsidies falling in-between those extreme cases, social welfare increases in some of the states and decreases in other demand states. In such a case, establishing the net welfare effect of the subsidy will require knowledge of the demand functions.

Parallel results on subsidies for the case when pollution distortions have extensive character are derived in the appendix D.1. Together, those results demonstrate the potential of generation subsidies for non-polluting resources to increase the efficiency of electricity markets.

The derivations for the theorem above focus on the efficiency effects on energy markets and assume the subsidy is paid from the general budget. However, in most real-life applications the subsidy is financed through revenue-neutral charges on electricity [Abrell et al. (2019)]. Under such financing structure, the net welfare effects of subsidies become non-trivial. The ambiguity occurs because of the asymmetric effects that subsidies have on the capacity and generation of non-subsidized resources (explained in Theorem 1 and Corollary 1) combined with the asymmetric effects that the electricity charge can have in different demand states and the heterogeneity in pollution intensity of non-subsidized resources. For instance, when the subsidized resources are just below relatively low-polluting generators in the merit order and when the demand curves are the steepest in demand states when those relatively low-polluting resources are on the margin, subsidies can cause pollution to increase, detracting from welfare.

Theorem 3. *When pollution is not internalized in the market, a sufficient condition for a subsidy recovered by a revenue-neutral consumption charge to increase welfare is that the status quo prices are distorted downwards in all demand states in which polluting resources are marginal.*

Proof - see appendix A.5. □

The intuition behind the theorem is that the consumption charge will work like a tax for polluting resources and like a subsidy for pollution-free resources. This is straightforward to see when prices are distorted in all states, for example, when there are no two consecutive demand states in which non-polluting resources are marginal. Here, any subsidy rate that leads to a sufficiently small price decrease in the states when pollution-free resources are marginal and a price increase lower than the smallest of price shortfalls in all the other states will increase welfare.

The sufficient condition from Theorem 3 is not met when a fossil-fueled generator type is much less polluting than the resource that follows it in the merit order. To see the effects of subsidies is such a setting, assume that there are three demand states and that the market is served by three types of resources: pollution-free baseload generators, low polluting mid-merit generators and highly polluting peakers. A subsidy combined with the consumption charge corrects p_1, increasing the capacity of the pollution-free resource at the expense of the low polluting generation, and corrects p_3 downwards. At the same time, however, it brings p_2 further away from optimal level. If, in such a setting, the peak demand and off-peak demand are highly inelastic while the intermediate demand state is very price elastic, a subsidy detracts from welfare – the additional volumetric charge causes a substantial shrinkage of the intermediate peak consumption but leaves the peak consumption almost unaffected. This causes the capacity of the most polluting resource, and thus pollution, to increase substantially.

Those findings contrast with the literature on two-part instruments which posits that it is always possible to reach an optimal solution in the presence of externalities through a combination of a subsidy for pollution-free energy producers and a consumption tax [Fullerton and Wolverton (2005)]. That difference in the results is due to heterogeneity in pollution intensity and the ensuing problem of having too many goals compared to available tools – an optimal pricing instrument would treat the different polluting resources differently, while a uniform subsidy for pollution-free generators cannot have a targeted, differentiated effect on the polluting generators.[23] Consequently, non-uniform generation subsidies that directly relate to the avoided pollution from the resource that the subsidized unit displaces are welfare dominant. In other words, the subsidized unit should not be rewarded for the absence of pollution, but rather for the avoidance of pollution.[24]

Proposition 3. *When there is heterogeneity in the pollution intensity of emitting generators, the optimal subsidy for a type of non-emitting generator should be specific to that resource and should account for the characteristics of the resources it displaces.*

[23]Some papers, for instance Goulder et al. (2016) and Eichner and Runkel (2014), show that subsidies for pollution-free resources can be more cost-effective than direct emission pricing or that feed-in-subsidies, in combination with other instruments, can help overcome the problem of lumpy entry costs [Antoniou and Strausz (2017)]. However, it is unclear to what extent those results hinge on the authors abstracting from different types of polluting generators and instead modeling only one type of emitting generator.

[24]Abrell et al. (2019) and Fell and Linn (2013) show how environmental values of pollution-free resources varies with the resources' intermittency profiles. Abrell et al. (2019) acknowledge that with such heterogeneity optimal subsidies should be differentiated by resource type.

A subsidy given to an i^{th}-merit resource affects only the i and $i-1$ prices. As a result, it changes the capacity of $i-1^{th}$-, i^{th}- and $i+1^{th}$-merit resources. Consequently, in accordance with Proposition 3, the optimal subsidy will depend on characteristics of those resource types only. □

These results have important insights for the design of clean energy policies, specifically renewable or clean energy standards. Driven by the insight that technology-neutral policy design leads to the most cost-effective abatement solutions, most jurisdictions use Renewable Energy Credits which provide uniform payments for all eligible technologies. Even when carve-outs for specific resources exist, those are usually motivated by usually supporting developing a nascent type of technology. However, our results are a reminder that technology-neutrality does not have to imply that all technologies end up receiving the same rate of subsidy. By coupling the payments to outcomes such as avoided pollution, regulators could still keep the payments technology neutral while maximizing their cost-effectiveness. If such an explicit coupling is not possible, developing new policy instruments such as zero-emission credits or offshore wind renewable energy credits might also be justifiable on economic theory grounds if those resources avoid different amounts of pollution.

3.4 Transition effects of subsidies

The derivations in Section 3.3 focus on equilibrium outcomes, which is also the prevailing approach taken in the literature [Joskow and Tirole (2007), Palmer et al. (2011), Briggs and Kleit (2013), Brown (2018a), Llobet and Padilla (2018), Bento et al. (2018) Özdemir et al. (2020)]. However, policymakers are often concerned with transition effects caused by the introduction of a policy.

In the energy sector, transition to a new equilibrium will tend to be protracted compared to, for instance, adjustment to new monetary policies,[25] making the intermediate effects more relevant. Given the paramount importance of reliability, policymakers will also avoid any policy that produces an adjustment path with temporarily lower grid reliability, even if that policy results in superior equilibrium outcomes. For instance, a reform resulting in misaligned timing of entry and retirement decisions could harm the reliability of the grid, and would thus be rejected by regulators.

Below, we delineate some of the short-run changes induced by introduction of subsidies

[25]For monetary policy, it is conceivable to have instantaneous adjustments [Auernheimer (1974)] while a transition under an energy sector regulation will unfold over multiple years at minimum. This happens as energy assets are very long-lived – some of the coal power plants that are still in operation are over 70 years old – and building generators require a substantial lead time. For instance, building a nuclear power plant takes at least 6 years, for natural gas the construction takes up to 24 months. Building transmission to interconnect a new generator to the grid can also take years.

to a market operating under the status quo. While our model does not allow us to solve for the exact transition pathway, we provide general observations on the changes in profitability of the generators and capacity adjustments.

Introducing a non-disruptive subsidy to an i^{th}-merit resource is, from the perspective of the subsidized unit, tantamount to a decrease in its marginal cost. However, the subsidy will lead to an immediate drop in some prices only if i^{th}-merit resource is marginal in more than one state. Otherwise, all prices initially remain the same. The increased profitability of i^{th}-merit technology induces new entry, ΔK_i, shifting out the part of the supply curve above the subsidized resource, such that prices in all states $j \geq i$ decrease.

This price drop will reduce the per MW profits of all generators that do not receive a subsidy:

$$\Delta \Pi_j = \sum_{k=j}^{N} \Delta p_k f_k \ , \tag{9}$$

whereby Π_j is the profit of non-subsidized j^{th}-merit technology and Δp_k denotes the short-time price change in state k in response to additional capacity of type i:

$$\Delta p_k = \begin{cases} \hat{D}_k^{-1}(\sum_{j=1}^{k} K_j) - \hat{D}_k^{-1}(\Delta K_i + \sum_{j=1}^{k} K_j) & \forall\, k \geq i \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Given that with constant returns to scale technologies equilibrium profits are zero, the drop in profits causes losses. Therefore, in the short-term, all non-subsidized resources receive exit signals.

The retirement will decrease some of the losses described by the Equation (9), and it will continue as long as there are unprofitable technologies. While we cannot predict the exact path of adjustments, we know that the equilibrium capacities and prices will be as described in Theorem 1 and Corollary 1. The effects for disruptive subsidies follow a parallel logic.

Our findings confirm that if a policymaker decides to introduce a subsidy, the profitability of all non-subsidized units will drop. However, this effect will be only temporary. After the offsetting retirements, the profits of generators will rebound.

4 Energy-plus-capacity markets

4.1 General findings

The previous Section demonstrates that when prices are allowed to fluctuate to ensure revenue adequacy, energy-only markets are sufficient for providing electricity reliably. However, many real-life electricity markets have a cap on energy prices, which restricts the functioning of the markets and gives justification to creating capacity markets.[26] Below we show consequences of a price cap, p^{max}, such that $p^E_{N-1} < p^{max} < p^E_N$.[27]

With capacity levels defined by Equations (3)-(4) but with peak price, p^E_N, replaced by p^{max}, each MW of capacity of each generator type would bears a loss of $(p^E_N - p^{max})f_N$. Retiring a fraction of their capacity would allow low- and mid-merit generators to recover some of the revenue lost: such retirement would increase p_1 and p_2 etc. and thus decrease the amount of "missing money" for remaining units of those types. For peakers, though, there is no similar mechanism that would allow the recovery of missing money. Consequently, a cap leads to under-investment in peak capacity, peak demand surpassing generation capacity, and, in turn, blackouts. To prevent that, market operators implement capacity markets.

Joskow and Tirole (2007) show that an energy market with a price cap p^{max} and a technology-neutral capacity market lead to the same equilibrium investment and generation as an energy-only market without a price cap if generators are price-takers and if, under the price-cap regime, consumers on real-time meters face an additional charge for every kWh they consume *in the peak period* equal to the capacity price.[28] In such a setting, in equilibrium, all generators submit bids into capacity markets corresponding to the the revenue they lose because of the cap. Therefore, their capacity market bid per MW, b_N, which also equals the capacity clearing price, is given by:

$$b^p = \pi^{loss} = (p^E_N - p^{max})f_N = (c_N - p^{max})f_N + I_N. \tag{11}$$

To see the outcomes equivalency, it suffices to note that consumers will face the same effective prices as before the introduction of the price cap as a result of the capacity price formation described above and the additional peak period charge.

In reality, utilities usually recover the per-MW capacity payments they pay, κ, through electricity consumption fee, λ, imposed on *each unit of electricity consumed*. Such recovery method causes the outcomes to diverge between energy-only and energy-plus-capacity markets, as visible when comparing the equilibrium allocations.

[26] For the maximum prices that energy can reach in the U.S. wholesale markets, see Figure 1 in Chang et al. (2018).

[27] We concentrate on price caps above p^E_{N-1} as with lower caps, capacity markets cannot restore the optimum investment (see the problem of instruments vs goals as described in Joskow and Tirole (2007)). With the current price caps implemented across the U.S., the assumption might not be empirically true. Nevertheless, it helps us focus on the mechanisms investigated.

[28] With stochastic outcomes, Mays et al. (2019) show that the existence of capacity markets tilts the resource mix towards the generator types with lower fixed costs and higher operating costs if risk trading is incomplete. Therefore, for the equivalence result, the extension of the framework to stochastic outcomes would require an assumption of complete risk trading.

To find the equilibrium under capacity payments recovered through consumption fee, we use our two-step procedure, noting that the set of economic technologies is not affected by capacity payment's financing, and neither is the merit order. The new peak price, p^c_N, that makes the peaker break even is now given by:

$$I_N + f_N c_N = \kappa + p^c_N f_N \;\; \Rightarrow \;\; p^c_N = \frac{I_N - \kappa}{f_N} + c_N = p^E_N - \frac{\kappa}{f_N}, \tag{12}$$

where the superscript c denotes outcomes under energy-and-capacity markets with consumption fee, and index superscript ^{E} stands for outcomes under energy-only design. An equilibrium capacity price, κ^c, leads to $p^c_N = p^{max}$ and is given by:

$$\frac{I_N - \kappa^c}{f_N} + c_N = p^{max} \;\; \Rightarrow \;\; \kappa^c = I_N - f_N(p^{max} - c_N). \tag{13}$$

To predict equilibrium price in demand state $N-1$, we search for price that makes the $N-1^{th}$ resource break even:

$$\begin{aligned} I_{N-1} + (f_{N-1} + f_{N-1})c_{N-1} =& \kappa + p^c_{N-1} f_{N-1} + f_N p^{max} \\ p^c_{N-1} =& \frac{I_{N-1} - I_N + c_{N-1}(f_{N-1} + f_N) - c_N f_N}{f_{N-1}} = p^E_{N-1} \end{aligned} \tag{14}$$

Generally, it can be shown that $p^c_i = p^E_i \;\; \forall i < n$ and $p^c_i = p^{max}$ for $i = N$. Consequently, the financing of capacity payment through consumption fee does not change the wholesale energy prices. This is intuitive given that generators need to break even and that price cap affects one demand state only.

Despite unchanged energy prices, the quantities consumed under the analyzed scenario deviate from quantities in energy-only markets. This happens because of the wedge equal to λ between consumer price and producer price in *all* demand states. With perfect foresight, λ just covers the capacity payments, i.e. $\lambda = \frac{D_N(p^c_N+\lambda)}{\sum_i D_i(p^c_i+\lambda)}\kappa$, such that we can obtain the equilibrium quantities by solving the following system of equations:

$$\begin{cases} D^c_N = & D_3(p^{max} + \frac{D_N(p^{max}+\lambda)}{\sum_j D_j(p^c_j+\lambda)}\kappa) \\ D^c_i = & D_i(p^c_i + \frac{D_N(p^{max}+\lambda)}{\sum_i D_i(p^c_i+\lambda)}\kappa) \;\; \forall i < N. \end{cases} \tag{15}$$

The resulting peak consumption is inefficiently high, leading to overbuilding of the peaker's capacity. For other resources, the direction of capacity change compared to energy-only market depends on the relative price sensitivity of the demand in the states in which they are marginal.

4.2 Comparing equilibrium outcomes under status quo and Pigouvian tax with energy-plus-capacity markets

As shown above, the effects of a capacity market on the resource mix depends on the exact design of capacity market. Absent a mechanism to internalize externalities, the market distortions with capacity markets are the same as the distortions under energy-only markets described in Section 3.2 when capacity payments are recovered from consumers through peak charges. With a consumption fee, on the other hand, the magnitude of the distortions might be weaker or stronger, depending on how the differences in capacity of various resource types implicitly defined by Equation (15) correlate with pollution intensity of those resources.

As both price formation and marginal emissions are the same under each demand state under energy-only and energy-plus-capacity designs, the Pigouvian tax leads to the same price adjustments under the two market designs.[29] However, the consumption effect of the tax differs between the designs when the demand curves are not isoelastic. This happens as the consumption fee, λ, causes the allocations to be on different part of the electricity demand curves.

Capacity prices are weakly lower under the status quo than under the first-best setting for any form of capacity payment recovery. Recall that capacity prices are determined by the difference between the competitive energy price in the highest demand state, which is the break-even point for the marginal generator, and the price cap. A Pigouvian tax would increase the effective marginal cost of the marginal generator if it is an emitting one, and hence the amount of revenue that is needed to break even, increasing the capacity prices.[30]

4.3 Equilibrium effects of subsidies with energy-plus-capacity markets

When studying the effects of generation subsidies on the outcomes in energy-plus-capacity markets, we again note that the existence of a capacity market does not change the equilibrium resource mix and the energy prices when capacity payments are recovered through peak charge. Therefore, the welfare impacts of subsidies in such a scenario are the same as described in Section 3.3. However, for recovery through consumption fee, some of the effects of subsidy change:

Corollary 2. *The effects of subsidies on energy prices are unaffected by the existence of*

[29] With the exception of peak price, which is capped under the energy-plus-capacity design.

[30] The peak price remains constant under the two regimes only if, under the status quo, a non-emitting generator as a peaker plant.

capacity markets as long as subsidies are not given to the peaker. Nevertheless, the consumption and resource mix effects of subsidies generally differ between energy-only market and energy-plus-capacity markets with a consumption fee.

It is easy to see that the equivalence in price formation process and in the final prices as shown in Equation 14 implies that the price effects of subsidies and Pigouvian taxes on prices also does not change between the two energy-plus-capacity market designs, no matter how capacity payments are recovered from consumers. Consequently, the findings from Theorem 1 remain unaffected. However, the magnitude of the consumption and investment effects of those instruments is different for energy-plus-capacity design as long as the price sensitivity of demand changes along the demand curve. This happens as the wedge in electricity prices, λ, causes that equilibrium allocations to shift along the electricity demand curve when producer prices (the wholesale prices) remain the same. □

Comparing prices under the status quo to prices with subsidies also leads to an important result relevant to current policy discussions in energy market discussions.

Corollary 3. *When giving a generation subsidy to a non-polluting resource does not change the identity of the generator type that is marginal in the peak period and this marginal generator is a polluting one, the subsidy does not affect the long-term competitive equilibrium capacity price.*

As shown in Theorem 1, subsidies have an effect on energy prices only in the states in which the subsidized resources are marginal and in the lower, adjacent demand states. As the long-term capacity price is determined solely by the uncapped peak price and the price cap, a subsidy that does not change the identity of the peaker and which is not received by a peaker will leave the capacity price unaffected. □

The intuition for capacity prices being unresponsive to generation subsidies relates to the fact that it is the energy market that re-equilibrates in response to generation subsidies. Capacities of the generators and energy prices adjust to new levels where all generators break even. For this finding, it is irrelevant here whether the subsidy is of disruptive or non-disruptive type.

Should the subsidies be paid to the peaker, the price p_N that the peaker needs to break even drops. This automatically reduces the "missing money" as described by Equation (11), thereby lowering capacity prices at the equilibrium. From the perspective of the non-subsidized resources, the drop in capacity prices is equivalent to a decrease in peak price: It would increase price in the demand state just below the peak demand, p_{N-1}, decreasing the capacity of technology marginal in that state and leaving all other prices unaffected.

Our results imply that the existence of capacity markets is irrelevant for the long-term welfare effects of subsidies if capacity markets are correctly designed. Consequently, in equilibrium, welfare implications of generation subsidies are like described in Sections 3.2 and D.1, implying existence of socially desirable subsidy designs. These results contradict the findings from previous studies on the impact of generation subsidies on the functioning of electricity markets, especially capacity markets. The main reason for the discrepancy is the fact that other studies ignored the existence of externalities, and therefore did not allow for the possibility that subsidies improve economic efficiency under certain circumstances.[31]

Further, our results contradict the conventional wisdom underlying the recently implemented policy reforms in wholesale capacity markets. Those reforms relied on the basic economic argument that subsidies would lead to price suppression in capacity markets, harming economic efficiency. While that argument would hold true in a simple market structure and when there are no externalities, the existence of capacity markets, the interdependent relation between energy and capacity markets, and externalities render the argument incorrect. The first-order effect of subsidies in the energy market implies that generation subsidies might lead to no change, or even an increase, in capacity market prices. In addition, the existence of externalities provide an additional channel for subsidies potentially increasing economic efficiency.

4.4 Transition effects of subsidies

Like in the case of energy-only markets, introduction of non-disruptive subsidy will at first have no effect on the energy market prices: While the subsidized technology will submit lower bids into the energy market, its existing capacity is limited, preventing lower energy prices. The subsequent rise in profit per MW of capacity for the subsidized technology will allow it to bid lower in the capacity market. However, the subsidized technology will not be marginal in the capacity market and thus, in the short run, it will not affect the capacity price.

As the subsidy will eventually attract new entry from the subsidized technology, the energy prices and profits of other generators, Π_j, will respond following the logic similar to that given in Equations (10) and (10). The losses of the non-subsidized resources will be reflected in the next auctions in their raised capacity bids:

$$b_j^p = \pi^{loss} + \Delta\Pi_j. \tag{16}$$

[31] Additionally, some papers include only one demand state, thereby not allowing the equilibrating processes like described in Theorem 1 to happen. Reliance on assumption of all generation types having the same marginal costs also precludes equilibrating mechanisms.

Since $\Delta\Pi_j$ differs between technologies, a sloping capacity supply curve forms, with the subsidized technology submitting the lowest bids.

Because the new supply curve in the short-term will be almost everywhere above the old curve, the capacity market will clear with a higher price. However, this, combined with the fact that total capacity in the market is now higher than before the introduction of the subsidy, implies that some capacity will fail to clear the market. The units without capacity market obligation will experience a stronger loss and will receive a market signal to exit. While it is hard to predict how exactly the retirement and investment decisions will be timed, the new equilibrium that the market will reach will be like that described in Section 4.3.

5 Application of the results to capacity market reforms in PJM and ISO-New England

In recent years, wholesale market operators in various regions of the U.S. have reformed their capacity market design. The justification of the reforms has been the need to shield the capacity markets from the much-feared possibility of "price suppression," which is claimed to follow from generation subsidies for non-polluting resources, such as Zero-Emission Credits and Renewable Energy Credits.

In March 2018, the Federal Energy Regulation Commission (FERC), which regulates wholesale energy markets in the U.S., accepted a new capacity market construct proposed by ISO-New England. The construct, called "Competitive Auctions with Sponsored Policy Resources," imposes a floor on the bid that subsidized new-generation resources can submit into the capacity market (FERC, 2018). This minimum bid is calculated to reflect the generator's costs, should the generator receive no subsidies.[32] In December 2019, FERC directed another wholesale market operator, PJM, to mitigate the capacity market impacts of subsidies by implementing a Minimum Offer Price Rule (MOPR). The MOPR rule prevents all subsidized resources from submitting bids into capacity markets lower than their unsubsidized costs (FERC, 2019). In another trading region, New York-ISO, a similar rule (referred to as Buyer Side Mitigation) is currently being extended in the context of subsidies for non-polluting generation (FERC, 2020).

[32] In other words, new, subsidized resources are allowed to submit bids corresponding to the difference between the energy market revenues the regulator expects them to earn and their private costs, but without taking into account the subsidies they receive.

These reforms have been controversial. Proponents argue that the reforms correct the price suppressive effects of subsidies and hence improve economic efficiency. Opponents argue that the changes harm pollution-free generators and hurt states' decarbonization efforts. The stakeholders used their own analyses to argue their point, and so far, no rigorous academic studies have emerged on the topic.

Capacity markets are a substantial source of revenue for generators, especially in case of peakers for which capacity revenue can be well over 50% of the total revenue of the generator. It is therefore important to understand how subsidies affect those outcomes and also how the capacity market reforms affect the functioning of subsidies. Our model framework allows us to do both, thereby enabling us to evaluate the justification for the reforms.

If, indeed, generation subsidies harm the economic efficiency of wholesale markets and MOPR-style policies prevent that effect from occurring, mitigation of the effects of generation subsidies should be undertaken in other trading regions as well. However, if such policies reduce social welfare, their implementation is not justified on economic efficiency grounds.

First consider the argument that subsidies lead to price suppression in capacity markets. As we show in Sections 3.2 and in the appendix D.1, energy prices in states when the subsidized resources are the marginal resources fall as a result of the subsidy due to the equilibrating process in the energy market. The price decrease continues to the point where, given the revenue from subsidies, the resource breaks even. Given this process, the equilibrium bids of subsidized resources in capacity markets continue to be defined by the "missing money" as described by Equation (11). In other words, the capacity bids of the subsidized resource types are not affected unless they are the peaker plant.

Consequently, if subsidies are given to low-merit order resources, as is currently the case for programs like Zero-Emission Credits or Renewable Energy Credits, the long-term equilibrium capacity bids of all resources that clear the energy markets stay the same with or without subsidies. In other words, in this setting, there is no long-time capacity price suppression effect from generation subsidies when they are given to non-marginal units.

The only cases in which subsidies decrease capacity prices would be when subsidies are given to the peaker or when subsidies change the type of technology serving as a peaker. In the period immediately following the introduction of such subsidies, the bids of the subsidized resource would decline compared to the equilibrium bid. At the same time, though, capacity bids of other resource types would (weakly) increase by $\Delta\Pi_j$ as defined in Equations (9) and (16) as subsidies for the peaker reduce peak energy market prices. As a consequence, the subsidized resource is no longer the price-setting unit in the capacity market and the capacity price (weakly) increases.[33] Our framework shows that "price suppression" in capacity markets does not occur even when the peaker plant receives a subsidy. The market reforms in this case are also not supported by our framework.

Overall, our results show that when the interactions between the energy and the capacity markets are taken into account, generation subsidies do not lead to price suppression in the capacity markets. On the contrary, we show that capacity prices might increase under certain circumstances. Hence, our framework shows that these capacity market reforms are not supported by economic theory.

Next, consider the effects of subsidies and capacity market reforms on the total social welfare. As we show in Section 3.3, there exists a range of subsidies that could bring the market closer to the socially efficient outcome. Assume a subsidy rate s^* has been introduced that enhances the efficiency of the market as discussed in Subsection 3.2. Assume also that the sector has transitioned to the new equilibrium given the subsidy.

In the short run, implementing MOPR could drive up the capacity prices towards $\pi^{loss,MOPR} = \pi^{loss} + s^* \sum_{k=i}^{N} f_k$, where k is the lowest demand state in which a subsidized resource is marginal. Such capacity prices would lead to extra profits for non-subsidized resources, and reshuffle the energy prices again, partly, through new entry of non-subsidized resources. This new entry, in turn, could prevent the subsidized resources from clearing the capacity markets. The aggregate costs of energy and capacity procurement would increase, reducing welfare. And, if the non-subsidized resources are also emitting resources, the social welfare shrink even further due to increased emissions. In other words, our framework shows that these reforms would lead to decreases in social welfare, a result that is the exact opposite of their intended effect.

As a result, we conclude that these recently implemented reforms, contrary to their intended effect, could lead to economic inefficiency when there are generation externalities. If the subsidies in questions directly address an externality and are within a certain range, they enhance social welfare. Instead, measures like MOPR should be considered only when a particular generation subsidy is found to be too high and thus welfare-decreasing.

6 Conclusions

Despite the mounting concerns over unsustainable greenhouse gas emissions and the consequent climate change, the political will to implement pollution taxes has been lacking, especially on national levels. Consequently, policymakers have embraced generation sub-

[33] A subsidy given to a resource type marginal in the lowest states would leave the capacity price unchanged in the short term and cause it to increase only after the subsidized unit starts expending its capacity.

sidies for non-polluting generators as a remedy. As the subsidy approach is increasing in importance, both in terms of the number of various subsidies used and in terms of aggregate magnitude of the payments, concerns have arisen about its impact on the functioning of wholesale electricity and capacity markets. In the U.S., two market operators, PJM and ISO-NE, are implementing reforms to mitigate the effects of subsidies on capacity prices. At the same time, academic literature on the total welfare implications of subsidies is scarce.

We contribute to this discussion by developing a framework in which the effects of subsidies on energy and capacity markets can be analyzed. We first confirm that not addressing externalities in wholesale markets skews the generation mix toward polluting resources. We then show how prices in energy and capacity markets respond to additional payments to non-polluting generators. In particular, we demonstrate that, as long as the subsidized resource is not the marginal resource in the peak period, generation subsidies do not affect the equilibrium price in capacity markets. We prove that, due to heterogeneity of polluting resources, a uniform subsidy cannot restore the first-best outcomes, even when combined with an energy consumption charge. A better subsidy design would compensate resources for "avoidance" of pollution instead of paying them for "absence" of pollution. Consequently, the potential for a subsidy to enhance welfare depends largely on how the relative pollution footprint of resources is distributed over the merit order. However, even a uniform subsidy can improve the efficiency of the markets when pollution is not internalized, which we argue by showing the existence of an efficiency-enhancing subsidy rate. As a consequence, policies that indiscriminately mitigate any subsidy, without taking its welfare effect into account, harm the economic efficiency of wholesale electricity markets. Our results show, based on economic theory, the recent major reforms in energy and capacity markets, which significantly alter the functioning of the markets, were fundamentally flawed.

References

Abrell, J., M. Kosch, and S. Rausch (2019). Carbon abatement with renewables: Evaluating wind and solar subsidies in Germany and Spain. *Journal of Public Economics 169*, 172–202.

Abrell, J., S. Rausch, and C. Streitberger (2019). The economics of renewable energy support. *Journal of Public Economics 176*, 94–117.

Acemoglu, D., P. Aghion, L. Bursztyn, and D. Hemous (2012). The environment and directed technical change. *American Economic Review 102*, 131–166.

Antoniou, F. and R. Strausz (2017). Feed-in subsidies, taxation, and inefficient entry. *Environ Resource Econ 67*, 925–940.

Auernheimer, L. (1974). The honest government's guide to the revenue from the creation of money. *Journal of Political Economy 82*, 598–606.

Bento, A. M., T. Garg, and D. Kane (2018). Emissions reductions or green booms? General equilibrium effects of a renewable portfolio standard. *Journal of Environmental Economics and Management 90*, 78–100.

Blumsack, S., C. Lo Prete, U. V. Shanbhag, and M. Webster (2018). Analysis of state policy interactions with electricity markets in the context of uneconomic existing resources: A critical assessment of the literature. Working Paper.

Briggs, R. and A. Kleit (2013). Resource adequacy reliability and the impacts of capacity subsidies in competitive electricity markets. *Energy Economics 40*, 297–305.

Brown, D. P. (2018a). Capacity payment mechanisms and investment incentives in restructured electricity markets. *Energy Economics 74*, 131–142.

Brown, D. P. (2018b). The effect of subsidized entry on capacity auctions and the long run resource adequacy of electricity markets. *Energy Economics 70*, 205–232.

Chang, J., M. G. Aydin, R. Broehm, Y. Yang, and R. Sweet (January 26, 2018). Shortage Pricing in North American Wholesale Electricity Markets. The Brattle Group Retrieved September 2020. `http://files.brattle.com/files/14169_4_3-brattle-paper-shortage-pricing.pdf`.

Eichner, T. and M. Runkel (2014). Subsidizing renewable energy under capital mobility. *Journal of Public Economics 117*, 50–59.

Federal Energy Regulatory Commission (FERC) (2018). 162 FERC ¶ 61,205 Order on Tariff Filing. March 9, 2018.

Federal Energy Regulatory Commission (FERC) (2019). 169 FERC ¶ 61,239 Order Establishing Just and Reasonable Rate. December 19, 2019.

Federal Energy Regulatory Commission (FERC) (2020). 172 FERC ¶ 61,058 Order Addressing Arguments Raised on Rehearing and Compliance. July 17, 2020.

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

Fischer, C. (2010). Renewable portfolio standards: When do they lower energy prices? *The Energy Journal 31*(1), 101–119.

Fullerton, D. and A. Wolverton (2005). The two-part instrument in a second-best world. *Journal of Public Economics 89*, 1961—-1975.

Goulder, L. H., M. A. Hafstead, and I. R. C. Williams (2016). General equilibrium impacts of a federal clean energy standard. *American Economic Journal: Economic Policy 8*, 186–218.

Haan, P. and M. Simmler (2018). Wind electricity subsidies - A windfall for landowners? Evidence from a feed-in tariff in Germany. *Journal of Public Economics 159*, 16–32.

Holland, S. P., J. E. Hughes, and C. R. Knittel (2009). Greenhouse gas reductions under low carbon fuel standards? *American Economic Journal: Economic Policy 1*, 106–46.

Independent Market Monitor for PJM (IMM) (2019). State of the Market Report for PJM 2018. Volume 2: Detailed Analysis.

Joskow, P. and J. Tirole (2007). Reliability and Competitive Electricity Markets. *RAND Journal of Economics 38*, 60–84.

Jung, C., K. Krutilla, and R. Boyd (1996). Incentives for Advanced Pollution Abatement Technology at the Industry Level: An Evaluation of Policy Alternatives. *Journal of Environmental Economics and Management 30*, 95–111.

Llobet, G. and J. Padilla (2018). Conventional Power Plants in Liberalized Electricity Markets with Renewable Entry. *Energy Journal 39*, 69–91.

Mays, J., D. P. Morton, and R. P. O'Neill (2019). Asymmetric risk and fuel neutrality in electricity capacity markets. *Nature Energy 4*, 948–956.

Oren, S., S. Smith, and R. Wilson (1985). Capacity Pricing. *Econometrica 53*, 545–566.

Özdemir, Ö., B. F. Hobbs, M. van Hout, and P. R. Koutstaal (2020). Capacity vs energy subsidies for promoting renewable investment: Benefits and costs for the EU power market. *Energy Policy 137*, 1–14.

Palmer, K., A. Paul, M. Woerman, and D. C. Steinberg (2011). Federal policies for renewable electricity: Impacts and interactions. *Energy Policy 39*, 3975–3991.

Reguant, M. (2019). The Efficiency and Sectoral Distributional Impacts of Large-Scale Renewable Energy Policies. *Journal of the Association of Environmental and Resource Economists 6*, S129–S168.

Steiner, P. O. (1957). Peak Loads and Efficient Pricing. *The Quarterly Journal of Economics 71*, 585–610.

Stoft, S. (2002). *Power System Economics. Designing Markets for Electricity.* IEEE Press & Wiley-Interscience.

A Proofs and derivations

A.1 Proof of Lemma 1

To prove that there exists a surjective mapping from the demand states onto the set of economic generators, we need to show that all economic generators are marginal in at least one demand state and that in each state at most one type of generation can join the merit.

For the first part of the proof, recall that each resource type has a unique marginal cost. This, combined with our competitive framework in which generators bid their true marginal costs in the auction, implies that no two bids are the same. Therefore, in each demand state there can be only one type of resource that sets the price and is thus marginal.[34]

For the second part, note that there cannot exist two i^{th}-merit of resources. Two i^{th}-merit resource types would receive the same revenue as they would clear the market in the same states but, by assumption, they would have different levelized costs of entry. Consequently, they cannot both be breaking even. As no generation type would be willing to incur losses in the long run, at least one type of the two resources must make profits. This would, however, induce new entry by the profitable resource type, thus reducing the prices until the level below which the other resource type makes loses and leaves the market.

We can thus conclude that in equilibrium, in state $i = 1$, only one type of resource provides electricity. As the demand increases (demand state switches from i to $i + 1$) at most one type of resource will join the group of resource types actively producing electricity. Consequently, we have that $m \leq N$. □

A.2 Proof of Lemma 2

Lemma 2 defines the marginal generator in state i as the generation type j for which the following inequality holds:

$$I_j + c_j \sum_{k=i}^{N} f_k \leq I_g + c_g \sum_{k=i}^{N} f_k \quad \forall g \in \mathcal{M}. \tag{17}$$

To show that this inequality indeed holds for i^{th}-merit generator, recall that the unit that is marginal in state i, clears also all states j where $j > i$. Consequently, the share of time that the i^{th}-merit generator is serving the market is $\sum_{k=i}^{N} f_k$, and the total costs associated with building 1MW of capacity of technology g and running it is given by the right-hand side of the inequality. In equilibrium, the technology for which those costs are the lowest becomes

[34] There might be two or more states that have the same marginal type of resource, e.g. when $M < N$.

the i^{th}-merit generator, since, given the sequence of prices $\{p_j\}_{j=i+1}^N$, which are independent of characteristics of i^{th}-merit generator, this technology is associated with the lowest p_i.

This inequality can also be seen by analyzing screening curves – curves plotting average cost of generation for individual technologies as a function of capacity factor.[35] □

A.3 Proof of Theorem 1

Recall that the peak equilibrium price is affected only by the cost characteristics of the N-merit generator, $p_N^E = c_N + \frac{I_N}{f_N}$. By definition, a non-disruptive subsidy will not change the identity of the peaker. Unless $i = N$, in which case the subsidy given to the peaker, the peak price remains unaltered. The equilibrium price in state $N-1$ depends on p_N^E and the characteristics of $(N-1)^{th}$-merit generator and thus does not change either, unless $i = N-1$. Similar logic applies to state $N-2$, $N-3$, and all other demand states k such that $k > i$.

As the subsidy effectively lowers the marginal costs of the i^{th}-merit generator by s, while the revenue per MW of capacity from all states $k > i$ remains constant, the rise in profits will attract new entrants of the i^{th}-merit generation type. Increasing capacity will suppress p_i to the point where new entry is not profitable anymore, i.e. $p_i^s = p_i^N - \Delta_i(s)$.

Given the decline in p_i, the price in state $i-1$ needs to rise by $\Delta_i(s)\frac{f_i}{f_{i-1}}$ for the $(i-1)^{th}$-merit generator to break even.[36] As the rise in p_{i-1} exactly compensates for the decline in p_i, the prices in states $k < i-1$ remain unaffected. □

A.4 Proof of Theorem 2

As we argue in Theorem 1, a subsidy received by an i^{th}-merit pollution-free resource decreases the clearing price in states i and raises the price in state $i-1$. To prove that those changes increase welfare, recall from Proposition 1 that, with a non-polluting i^{th}-merit resource, the price p_i is inflated whenever $(i+1)^{th}$-merit generator is polluting. At the same time, the price in state $(i-1)$ is below its optimal value whenever the marginal generator in that state is a polluting one. Consequently, a small enough generation subsidy brings prices closer to their optimal value whenever the technologies adjacent in the merit order

[35] For the logic of screening curves see Stoft (2002) and "efficiency ranges" in Oren et al. (1985).

[36] Note that we are looking at subsidies small enough to be non-disruptive. This, combined with the observation that with a competitive environment the resources just breaks even without the subsidy, guarantees that the subsidy does not decrease the price below marginal costs, i.e. $p_i - \Delta_i^s > c_i - s$.

are polluting. For instance, all subsidies in the range defined by:

$$\{s \in \mathbb{R}_+ \,|\, (\max_{j \in \mathcal{P}_{f,clean}} -\Delta_j(s) < \min_{j \in \mathcal{P}_{f,clean}} p_j^{\Delta}) \;\wedge\; (\max_{j \in \mathcal{P}_{f,dirty}} \Delta_{j,-1}(s) < \min_{j \in \mathcal{P}_{f,dirty}} -p_{j-1}^{\Delta})\},$$

where $\mathcal{P}_{f,clean}$ denotes the set of non-polluting economic resources, bring all prices they affect closer to their optimal value and thus unambiguously increase market efficiency.[37]

What happens if two or more pollution-free resource types follow each other in the merit order, e.g. when both i^{th}- and $(i-1)^{th}$-merit resources are pollution free, such that the competitive price p_{i-1}^E coincides with the optimal price p_{i-1}^E? In such a situation, the welfare change associated with the effect of a subsidy on p_{i-1} is negligible – when evaluated at the optimal price, the derivative of welfare with respect to price equals zero. Consequently, the subsidy will have a positive impact on market efficiency even if pollution-free resources follow each other in the merit order. □

A.5 Proof of Theorem 3

Assume for simplicity that the charge to finance subsidy, $t(s)$, is paid by generators for each MWh they produce. The revenue neutrality implies that the charge needs to meet the following condition:

$$t(s) \sum_{i \in \mathcal{M}} \hat{D}_i(p_i(s)) = s \sum_{j \in \mathcal{P}_{f,clean}} K_j(s) \sum_{k=j}^{N} f_k.$$

Any revenue-neutral subsidy in the set defined by:

$$\{s \in \mathbb{R}_+ \,|\, \max_{j \in \mathcal{P}_{f,clean}} -\Delta_j(s) - t(s) < \min_{j \in \mathcal{P}_{f,clean}} p_j^{\Delta} \\ \wedge \;\; t(s) + \max_{j \in \mathcal{P}_{f,dirty}} \Delta_{j,-1}(s) < \min_{j \in \mathcal{P}_{f,dirty}} -p_{j-1}^{\Delta}\}, \tag{18}$$

brings each of the wholesale prices closer to their optimal values as it leads to a sufficiently small price decrease in states when pollution-free resources are marginal and a price increase lower than the smallest of price shortfalls in all other states.[38]

If pollution-free resources are marginal in two or more consecutive demand states, e.g. in i and $i-1$, the welfare effect of a price change will be negligible for state i. This implies that the subsidy identified by condition (18) will have a positive impact on market efficiency

[37] The above set of subsidies is not empty since $\min_{j \in \mathcal{P}_{f,clean}} p_j^{\Delta}$ and $\min_{j \in \mathcal{P}_{f,dirty}} -p_{j-1}^{\Delta}$ are fixed, positive numbers, while $\|\Delta_j(s)\|$ can be set arbitrarily small by decreasing s.

[38] By definition, $\Delta_j(s) < 0$ and $\|\Delta_j(s)\| > t(s)$ whenever j-merit resource is of pollution-free type.

independent of relative location of pollution-free resources in the merit order. □

A.6 Proof of Theorem 4

To see that there exists a weakly welfare-increasing subsidy financed from a general budget, assume first that no pollution-free resource types are economic under the status quo but at least one of them is economic under first-best outcomes. Denote by $\mathcal{D}$ the set of demand states under which pollution-free resources are marginal under first-best. In such a case, the generation subsidy for pollution-free resources of rate

$$s = \min_{j \in \mathcal{D}} \{s_j^{min}\} \tag{19}$$

is guaranteed to increase welfare as it corresponds to the lowest of minimum subsidies s^{min} for all pollution-free resources that are inappropriately included in the merit order.

If, on the other hand, at least one pollution-free resource type belongs to the set of economic resources under the status quo, $\mathcal{P}_N^{SQ}$, a subsidy defined by (19) might be too high. However, by Theorem 2, we can always find a subsidy rate that reduces the distortion on the intensive margin in a way that increases market efficiency. □

B Results including consumers on traditional meters

B.1 Optimal capacity charges for consumers

A simultaneous introduction of price cap $p_{N-1}^E < p^{max}$ and of a capacity market leads to the same economic outcomes as an electricity market without a price cap if generators are price-takers and if under the price-cap regime consumers on the real-time meters face a charge for every kWh they consume in the peak period, PC, equal to the capacity price, while an amount of $\gamma = \textit{capacity price} \cdot \frac{f_N D_N'}{\sum_{i=1}^N f_i D_i'}$ is added to the price p paid by consumer on traditional meters.

For consumers on traditional meters with constant price sensitivity, the adder γ simplifies to $\pi^{loss} f_N$.

B.2 Price distortions with energy-only markets and no merit-order change

What is the magnitude of price distortion p^{Δ} for consumers on traditional meters when externalities are not accounted for? Formula 5 describes the optimum price p^* as a weighted average of the (optimal) wholesale prices. If the utility decides to follow that formula using the wholesale prices that are not optimal (i.e. where the externalities are not accounted for), the distortion associated with energy consumption by consumers on traditional meters becomes:

$$p^{\Delta} = p^{SQ} - p^* = \frac{e_1 D_1' + \sum_{i=2}^{N} e_i (D_i' - D_{i-1}') \sum_{k=i}^{N} f_k}{\sum_i f_i D_i'}. \tag{20}$$

It thus depends on the magnitudes of the externalities and the price sensitivity of demand.[39]

On the other hand, if the utility computes the price for consumers on traditional meters based on the ideal wholesale markets and not on the observed ones, there will be no price distortion compared to the first-best.

C Energy-only market with continuum of states of nature

The following model incorporates a continuum of states of nature like in the benchmark model of Joskow and Tirole (2007). The notation follows the main body of the paper with the addiction of $u_i(j)$ denoting utilization rates of plant of type j in state i, with $u_i(j) \in [0,1]$.

Social planner chooses prices for consumers on real-time and traditional meters, $\hat{p}_i$ and p respectively, the interruptibility parameters, α_i, the utilization rates and the amount of investment in each technology, $k(j)$ to maximize welfare:

$$\begin{aligned} \max_{p,\hat{p}_i,\alpha_i,u_i(c),K()} W = & \int_0^1 S_i(p,\alpha_i) + \hat{S}_i(\hat{p}_i) f_i \, \mathrm{d}i - \int_0^{\infty} \int_0^1 (c_j + \mathbf{e_j}) u_i(j) f_i \, \mathrm{d}i \, dK(j) \\ & - \int_0^{\infty} I_j dK(j) \\ & \text{s.t. } D_i(p,\alpha_i) + \hat{D}_i(\hat{p}_i) = \int_0^{\infty} u_i(j) dK(j)) \ \forall i. \end{aligned}$$

The associated first order conditions are as follows:

$$\mathbb{E}_i[\frac{\partial S_i}{\partial p} - \frac{\partial D_i}{\partial p}\frac{\lambda_i}{f_i}] = 0 \qquad \text{(F.O.C. wrt } p\text{)}$$

[39]Notice that if the slope of demand of the consumers on traditional meters does not change here across the demand states and wind is the lowest-merit resource, i.e. $e_1 = 0$, all terms in p^{Δ} cancel out in which case those consumers see the first-best prices. This is driven by the fact that the optimum prices account for the pollution content of the resource that is next in merit order and by the lowest merit-order resource being a pollution-free one. When the latter is not true, the distortion of p is proportional to e_1.

$$(c_j + \mathbf{e_j}))f_i = \lambda_i \qquad \text{(F.O.C. wrt } u_i(\cdot))$$

$$I(c) = \mathbb{E}_{i:\lambda_i/f_i > c_j + e_j)} \left[\frac{\lambda_i}{f_i} - c_j - \mathbf{e_j} \right] \qquad \text{(F.O.C. wrt } k(\cdot))$$

$$\frac{\frac{\partial S_i}{\partial \alpha_i}}{\frac{\partial D_i}{\partial \alpha_i}} = \frac{\lambda_i}{f_i} \qquad \text{(F.O.C. wrt } \alpha_i)$$

$$\lambda_i = \hat{p}_i f_i \qquad \text{(F.O.C. wrt } \hat{p}_i)$$

$$\int_0^\infty u_i(j) dK(j) = D_i(p, \alpha_i) + \hat{D}_i(\hat{p}_i), \qquad \text{(F.O.C. wrt. } \lambda_i)$$

where λ_i is the Lagrange multiplier associated with the need to have sufficient energy generation to cover the aggregate demand in state i. The above conditions determine the solution to the problem. In the optimum, the consumers on smart meters face the real-time wholesale energy prices and the traditional-meter consumers are charged a price that is a weighted average of the wholesale energy prices.

Importantly, pollution externalities affect both the optimal amount of capacity and the utilization rates of individual types of generators (the new elements compared to the model without environmental externalities are highlighted in bold). The direction and magnitude of that influence become clear when looking at closed form solutions, which we obtain when restricting the number of states of nature and generator types to be finite, as done in Section 3.3.

D Welfare comparisons when externalities create extensive distortions

The inefficiencies associated with not internalizing externalities are particularly pronounced when the externalities are disruptive, i.e. when they change the merit order defined by the mapping $h : \mathcal{N} \mapsto \mathcal{M}$. In such a case, the welfare losses under the status quo occur through two channels:

- distortion on the intensive margin where, for some of the consumed energy units, the social marginal costs of generation are higher than the utility from consumption
- extensive distortions where, in some of the demand states, the marginal technology is

not the least expensive one from a social point of view, resulting in unnecessarily high social generation costs given the quantities produced.

D.1 Comparison of equilibrium outcomes in energy-only markets under status quo, emission tax and subsidies regime

The first-best generation mix differs from status quo generation not only in the shares of individual resource types but also in the merit order. The latter implies that the first-best set of economic resources, $\mathcal{P}_N^{FB}$, includes additional types of the relatively low polluting resource types or that the low polluting resources are marginal in more states under the first-best. Correspondingly, under first-best, some types of higher polluting resources that are in the merit order under the status quo are not economic or they have lower capacity utilization.

Given the high efficiency losses that occur when merit order is distorted by the markets not accounting for externalities, the potential gains from a subsidy are higher. However, unlike with subsidies aiming at intensive margin presented in the previous subsection, there exists a minimum subsidy rate, s^{min}, below which the subsidy is ineffective – it does not change the merit order.

To see that, assume that the status quo merit order consist of various polluting resources with the i^{th}-merit resource being of type q. However, under first-best, a pollution-free resource of type r is marginal in state i. In such a case, in accordance with Lemma 2, the minimum subsidy rate needed to correct the merit order is given by:

$$s_i^{min} = c_r - c_q + \frac{I_r - I_q}{\sum_{k=i}^{N} f_k} + \epsilon, \tag{21}$$

where ϵ is an arbitrarily small positive number. Under such subsidy, the prices and demands in individual states are the same as without the subsidy, but the technology q gets (partly) replaced by resources of type r.[40]

Obviously, a uniform subsidy is again unable to fix the distortions in the merit order when the inefficiency concerns two resources of differing pollution intensity.

Theorem 4. *When the merit order under the status quo deviates from the first-best merit order because of uninternalized pollution, there exist a subsidy for non-polluting resources financed from the general budget that weakly increases efficiency of the market compared to the status quo.*

[40]The replacement might not be full if technology q is marginal under the status quo in more than one state.

Proof - see appendix A.6. □

When the subsidy is financed through a revenue-neutral electricity consumption charge, conditions parallel to those described in Theorem 3 are sufficient (but necessary) to ensure that the subsidy increases welfare.

D.2 Comparison of equilibrium outcomes in energy and capacity markets under status quo, emission tax and subsidies regime

As explained in Sections 4.2 and 4.3, an energy price cap combined with a capacity market can lead to the same outcomes and have the same welfare properties as an energy-only market. Therefore, the insights from energy-only markets on energy prices, resource mix and welfare under the status quo and generation subsidy in presence of extensive externalities (shown in Appendix D.1) are true also for energy and capacity market design.

Consequently, even with extensive externalities, the capacity prices are weakly lower under the status quo than under the first-best. This happens because capacity prices are determined by the difference between the competitive energy price in the highest demand state and the price cap. Emission pricing in the presence of extensive externalities increases the market peak price, unless, under the status quo, a pollution-free generator serves as a peaker.

The results on capacity prices in the setting of intensive externalities extend to the extensive externalities case. The capacity price under the status quo is the same as that under the subsidy scenario as long as the peaker unit under the subsidy regime is not of pollution-free type (and thus does not receive payments). This happens when the status quo peaker unit is of polluting type and the generation subsidy does not cause it to be replaced by a pollution-free unit. Should this not be the case, the long-term capacity price goes down.