Long-Run Equilibrium Modeling of Alternative Emissions Allowance Allocation Systems in Electric Power Markets*

Jinye Z. Schulkin†, Benjamin F. Hobbs‡, and Jong-Shi Pang§

September 13, 2007

Abstract

A question in the design of carbon dioxide trading systems is how allowances are to be initially allocated: by auction, by giving away fixed amounts, or by allocating based on output, fuel, or other decisions. The latter system can bias investment, operations, and pricing decisions, and increase costs relative to other systems. A nonlinear complementarity model is used to investigate long-run equilibria that would result under alternative systems for power markets characterized by time varying demand and multiple generation technologies. Existence of equilibria is shown under mild conditions. Solutions show that allocating allowances to new capacity based on fuel use or generator type can distort generation mixes, invert the operating order of power plants, and inflate consumer costs. The distortions can be smaller for tighter CO2 restrictions, and are somewhat mitigated if there are also electricity capacity markets or minimum-run restrictions on coal plants.

Subject classification: emissions trading, allowance allocations, electricity, air pollution, auction, grandfathering, cost-effectiveness, greenhouse gases, climate change, global warming, carbon dioxide, generation investment

JEL Classification Numbers: C61; L94; Q4; Q53

1 Introduction

Pollution cap-and-trade policies operate by allocating or selling permits to emit (or "allowances") to eligible pollution sources, who are then allowed to trade permits among themselves so that every polluter holds a number of allowances at least equal to their emissions. If the cap allows fewer emissions than would otherwise occur, the allowances have a positive market price. Under certain assumptions, such systems result in least-cost control of the emissions [24].

*The work of the first and third author was based on research supported by the National Science Foundation under grant CMMI-0516023. The work of the second author was supported by the National Science Foundation under grant ECS-0621920 and by the Energy research Centre of the Netherlands (ECN). The authors thank J. Sijm of ECN, C. Norman of JHU, and K. Neuhoff of Cambridge University for valuable suggestions.

†Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180-1590, U.S.A. Email: zhaoj2@rpi.edu.

‡Department of Geography & Environmental Engineering and Department of Applied Mathematics & Statistics, The Johns Hopkins University, Baltimore, Maryland 21218, U.S.A. and Scientific Advisor, ECN, Policy Studies Unit, Amsterdam, The Netherlands. Email: bhobbs@jhu.edu.

§Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, IL, USA. Email:jspang@uiuc.edu. This paper was written when the author was at the Department of Mathematical Sciences and Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, Troy, New York 12180-1590, U.S.A.

The first large-scale emissions cap-and-trade system was instituted in the US for electric sector SO_2 emissions by the 1990 Clean Air Act Amendments. Since then, cap-and-trade systems have been adopted in the US for NO_x and proposed by the US Environmental Protection Agency for mercury. Meanwhile, the EU has leapfrogged the US by adopting a cap-and-trade policy for the greenhouse gas CO_2. This system, called the Emissions Trading System (ETS), came into effect in 2005. Meanwhile, although there is no US federal CO_2 reduction requirement as of 2007, several states have initiated their own emissions reduction efforts, the most noteworthy being the Regional Greenhouse Gas Initiative (RGGI). CO_2 cap-and-trade programs are anticipated to have much larger economic impacts than previous emissions trading programs. The cost involved with significant CO_2 reductions and the economic value of trading such allowances are likely to be about an order of magnitude greater than for NO_x and SO_2 [9]. Wholesale electricity price increases in Germany and the Netherlands of 40% or more in 2005 have been blamed upon the introduction of the ETS [21].

There are many aspects of the design of a cap-and-trade system that can affect its economic efficiency and impacts upon consumers. One of the most important–and most debated–features is the initial allocation of allowances. The potentially high value of CO_2 allowances means, for example, that tens of billions of dollars of economic rents are created by the ETS system. Understandably, the power industry would prefer that this rent be given to them through a free initial assignment of allowances to existing and perhaps new power sources, while others argue that the government should auction the allowances and use the resulting revenues for tax relief or public programs [12]. However, not only income distribution is affected by who gets the economic rents associated with CO_2 allowances. Rules for initial distribution of allowances can also affect economic efficiency, potentially distorting investment, operation, and output pricing decisions and raising the social cost of reducing emissions [20]. "Social cost" is defined here in the manner usually used by economists as the sum of producer and consumer surpluses. We do not consider external pollution costs. As an example of such a distortion, if a contingent allocation system (also called "output-based" or "input-based" allocation) gives allowances in a way that depends on present or future generator decisions, incentives to deviate from least-cost investment mixes and operation are introduced. For instance, in the EU ETS, allowance allocations after 2007 depend upon emissions in 2005-2007, arguably providing an incentive to expand pollution in these years. The potential for distortion is clear from the analysis in [1] showing that the value of ETS allowances can be 70–105% of fixed plant costs. As another example, if a present polluting facility would lose its allowance allocation if it shuts down, then this provides an incentive to keep non-economic capacity in operation. On the other hand, if new investment is allocated free allowances, this can instead create a bias towards new investment. Further if dirtier new sources are allocated more allowances per unit capacity or unit output, as is done in at least eight EU countries [15], then technology choices can be skewed. Also, if different jurisdictions in the same power market have different allocation rules, as in the EU ETS, the location of investment can also be distorted.

Economic inefficiencies can result not only from distortions in production, but also from distortions in product pricing and consumption. As an example, free allocation to new entry can distort overall market prices, depressing prices below socially optimal levels if entry is made artificially cheap by the free provision of allowances. Also, existing distortions in retail power prices arising from average-cost based price regulation, which is still the rule in many US states and some EU countries, can be worsened by free allowance allocation [3]. There are strong political forces that support contingent allocation schemes, and the result is that most of the national allocation plans in the EU ETS are based upon such schemes [17, 23]. This support is in spite of numerous modeling studies that have compared the relative efficiency of such schemes compared to grandfathering and auctions, and that have found the latter to be superior.

In general, theoretical analyses show that allocating allowances in proportion to output tends to result in greater sectoral output (since there is an implicit subsidy of output), greater reductions in emissions rates per unit output, and higher control costs than grandfathering or auctioning, if there are not other

distortions in the economy [7]. However, in the presence of other market failures, such as inefficient tax policies or emissions policies that differ among sectors or countries, output-based policies can actually be welfare improving relative to grandfathering and auctions [5, 8]. As an example, if the cement industry is subject to CO_2 limits in some countries but not others, and cement is internationally traded, an output-based allocation in the regulated countries can lead to less distortion and lower social cost [4].

In this paper, however, we focus on the effects of allocation policies on a single market sector (power) that we assume is subject to the same rules throughout the market. We propose models for evaluating the long-run implications of different emissions allocation schemes for economic efficiency and consumer costs of the electric power sector. We compare investment, operating, and pricing outcomes of two general allocation approaches: contingent allocation schemes that allocate allowances free to new investment, and systems in which the initial allocation of allowances does not depend on present or future capital or operating decisions (either grandfathering or auction). The models can be used to investigate whether statements such as the following are likely to be true: "If the expansion of the generation park (by incumbents or newcomers) is associated with a free allocation of emission allowances, then players will base their long-term investment decisions on the long-term marginal costs, including the costs of the CO_2 allowances, but by subtracting the subsidy that lowers the required mark-up for the fixed costs. ... On balance, the power price will not be increased (ceteris paribus)" [14]. We do not address other important issues concerning the design of emissions allocation systems. Some of these include [20]: transparency and transactions costs; international competitiveness of affected industries; which economic sectors are covered; possibilities for obtaining allowances by funding emissions reductions in developing countries; the effect of mechanisms, such as price ceilings, designed to stabilize prices; the value of "banking" schemes to buffer interannual variations in emissions; or the efficiency implications of different ways to dispose of auction revenues.

Previous modeling studies of the power sector can be divided into two groups. The first includes detailed simulation analyses of near term (e.g., 2005–2025) market developments using large-scale linear programming or other optimization-based models for calculating market equilibria. The second consists of theoretical analyses designed to show general results, often for the long-term. Short-run analyses have considered the present mix of generation capacity in particular markets, and simulated competitive entry of new generation over the next decade or two under alternative allowance schemes. For instance, Neuhoff, Grubb, and Keats [16] use the IPM linear programming model to simulate effects on coal and natural gas investments, prices, and generator revenues under an exogenous (fixed) CO_2 price and no demand elasticity. As another example, Bartels and Musgens [2] apply a linear programming model formulated for 11 European power markets, and find that giving all new capacity the same number of allowances irrespective of emission rates ("sector benchmarking") resulted in less distortion than fuel-specific formulas that gave more allowances to technologies with more emissions. More coal plants were added in the latter case than under either sector benchmarking or allowance auctions. An earlier study [3] applies *Haiku*, a multidecadal equilibrium model for the US. The latter model, unlike the above linear programs, considers price elasticity and average cost-based regulation of retail electricity prices. As a result of inefficient retail pricing, grandfathering is found to have much higher social costs than auctioning, unlike in other studies. Finally, Palmer, Burtraw, and Kahn [18] use *Haiku* to determine the minimum fraction of allowances that should be given away to generators in order to ensure that they would not be worse off after the implementation of allowance trading; this number (approximately 20% for the RGGI program) was surprisingly small.

In contrast to these studies, theoretical analyses tend to involve simpler models and more general conditions. Some theoretical analyses use two-period models to address the distortion that arises if decisions in one period affect emissions allocations in the next period, as in the first two phases of the EU ETS. These demonstrate the existence of a bias towards over-investment in the first period [2, 16, 23] to gain more allowances later. But Neuhoff, Martinez, and Sato [17] point out that incentives to new entry can help mitigate market power in existing concentrated markets, and can also offset a

bias towards keeping old plants running if shutting them down would cause allowances to be forfeited. Other theoretical results include the following. If allowances are given free to new investment, this increases the effective demand for allowances by generators, inflating the price of emissions allowances and the cost of compliance if power demands are fixed [23]. The papers [16, 17] explore how long-run choices between two new generation technologies could be distorted by fuel-specific allocation rules compared to auctions assuming a zero-profit, free-entry equilibrium. They show that distortions are worse if the price of allowances is fixed, as dirty technologies will significantly expand if given more allowances than clean technologies. However, the investment distortions are less (but power prices are higher) if instead emissions are capped, because emissions prices rise. An innovative long run analysis by Smeers and Ehrenmann [22] looks at the complications introduced by particular market failure in the power market: the existence of market power mitigation rules in the power market that can depress returns on investment and yield suboptimal capacity additions. In a second-best analysis, they show that it is possible to design a free allocation of allowances to new generation capacity that can largely offset those investment disincentives, and actually improve market efficiency.

The models of this paper are more elaborate than other theoretical analysis of allowance allocation in power markets, with the possible exception of [22] because of its consideration of capacity market failures. The most comparable models are those of [16, 17]. Like those models, we consider a long run, free-entry equilibrium among more than one technology. We also share their implicit assumptions that long-run contract markets and short-run spot markets are arbitrage, that generators are price takers (although firms exhibiting market power can be handled easily), and that there are constant returns to scale in generation. The models of this paper are, however, more general than previous theoretical models in the following respects. Neuhoff, Grubb and Keats [16] consider up to two supply technologies, and assume a fixed operating order; in particular, when they consider two technologies, they assume that coal plants are always operated in preference to gas plants. Our model is more general in that any number of plant types can be considered, and the operating ("dispatch") order is endogenous; this is important, because our solutions show that dispatch orders can change if allowance prices are high enough. Our model also automatically considers corner solutions, in which some plant variables are zero. Also, minimum output constraints can be imposed, reflecting the reality that some types of capacity (modern coal plants) cannot be cycled on and off on a daily basis; this results in a more realistic characterization of the ability of power supply systems to adapt to changing emissions prices. Capacity markets are included in addition to energy markets, unlike other models. On the demand side, our model can consider arbitrary temporal distributions of demand, which can be price-responsive, unlike the models of Neuhoff and his colleagues. Finally, our models allow the number of allowances allocated per MW of new capacity (capacity-based allocation) or per MWh of energy output (sale-based allocation) to be endogenous. In particular, the allocation rule specifies the total number of allowances that are to be available to new capacity, and the amounts per MW or per MWh are automatically adjusted to achieve that target. Since several EU national allocation plans place a ceiling on the number of allowances to be allocated to new investment, some type of rationing similar to this may need to be instituted when enough entry has occurred so that ceiling is reached [20]. However, a price is paid for this added complexity; our general models are formulated as nonlinear complementarity problems for which analytical solutions cannot be derived. This means that it is not possible to obtain general analytical results showing the equilibrium as an explicit function of parameters, unlike [2, 16, 23]. Instead, our models need to be solved repeatedly for different parameter sets. Further, the inclusion of endogenous per MW or per MWh allowance allocations results in bilinear equilibrium conditions that make it more difficult to compute or show the existence of equilibria.

The paper is organized as follows. Following a statement of the model, along with several variants, we demonstrate that a solution exists under general conditions. By using PATH solver, the model is solved under several sets of input assumptions in order to explore the inefficiencies that result from different emissions allocation systems. In particular, we calculate the inefficiency that results from two different rules for allocating allowances to new investment, both of which discriminate between different plant

technologies: one that allocates allowances based on a per MW of capacity rule (potential emission rule), and another that allocates allowances in proportion to emissions (actual emission rule). The results show that as the percentage of allowances granted free to new construction increases, the inefficiency (quantified as the loss of producer and consumer surplus) also increases. We also consider how the results are affected by the simultaneous imposition of a capacity market, as well as by the presence of minimum-run constraints that more realistically simulate the operation of coal plants.

2 Model Definition

We summarize the notation used in the model formulation; first the parameters, which are all nonnegative, next the input functions, and finally the models' variables, which include the firms' variables and the market prices of capacity and emission allowances. The physical units are noted within parentheses. Figure 1 depicts the various components in the market structure and their interrelations.

Parameters: all positive except possibly CAP_f and $\underline{\mathrm{CAP}}$ which can be zero,

$\mathcal{F}$	Set of firms
$\mathcal{T}$	Set of time periods $\equiv \{1, \cdots, T\}$
CAP_f	Minimal amount of energy that firm f has to generate (MW)
MC_f	Marginal cost for firm f, excluding cost of emission allowances (EURO/MWh)
E_f	Emission rate for firm f (tons/MWh)
F_f	Annualized investment cost of firm f's capacity (EURO/MWyr)
R_f	Ratio of allowances allocated to firm f per unit of capacity relative to firm 1
$\widehat{R}_f$	Ratio of allowances allocated to firm f per weighted unit of sales relative to firm 1
$\overline{E}$	Total emission allowances supply (tons/yr): $\overline{E} > E_{\mathrm{GF}}$
E_{GF}	Amount of emission allowances that are grandfathered or auctioned (tons/yr)
H_t	Hours in period t (hr/yr), here assumed to be $8760/T$
χ	Unit converter = 1 MW^2 yr/EURO
$\underline{\mathrm{CAP}}$	Total capacity requirement (MW)

$\overline{E} > E_{\mathrm{GF}}$ means that a certain volume of free allowances is guaranteed to new entrants. For example, In EU ETS phase I, a fraction of the allowances have been reserved for eligible new entrants.

Functions:

$d_t(\bullet)$	Demand function for energy, strictly decreasing (MW)
$\pi_t(\bullet)$	The inverse of $d_t(\bullet)$; (EURO/MWh)
$e_{NP}(\bullet)$	Nonpower emission, nonincreasing (tons/yr)

Variables:

p_t	$= \pi_t\left(\sum_{g\in\mathcal{F}} s_{gt}\right)$: Energy price during period t (EURO/MWh)
p_e	Emission allowance price (EURO/ton)
p_c	Capacity price(EURO/MWyr)
α_f	Emission allowance for firm f (tons/MWyr)
s_{ft}	Energy sold by firm f in period t (MW)
$\bar{s}_{ft}$	$= s_{ft} - \mathrm{CAP}_f$ (MW)
cap_f	Capacity for firm f (MW)
μ_{ft}	Dual variable associated with firm f's capacity constraint in period t (EURO/MWyr)

There are three main components in the basic model: (a) firms' profit maximization problems, (b)

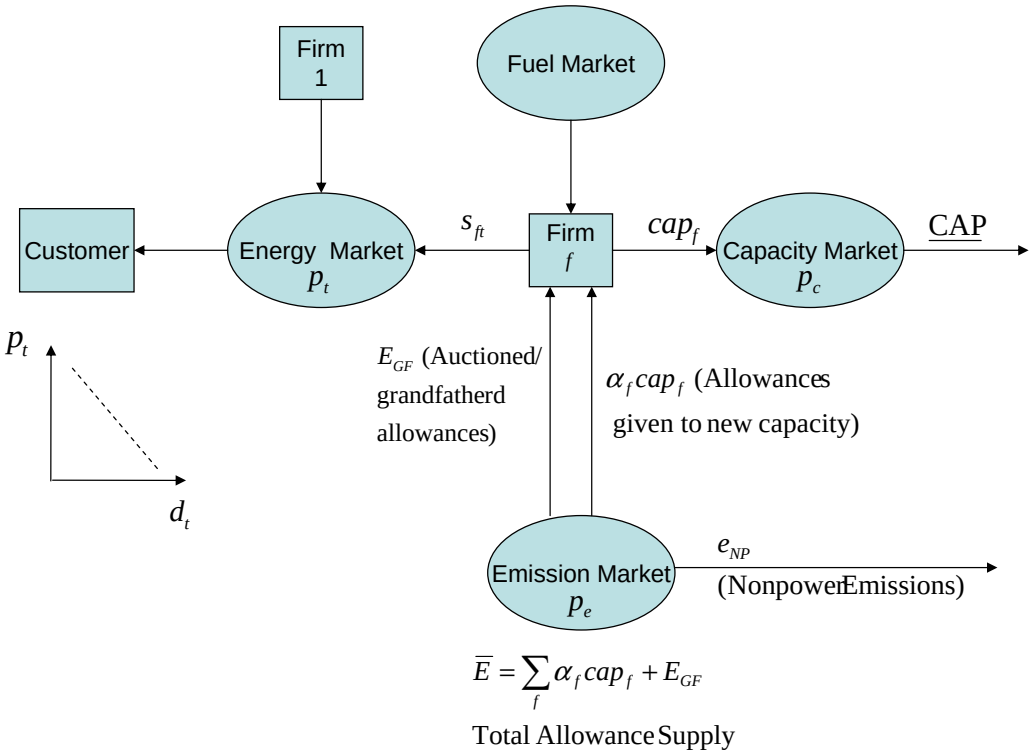

Figure 1: Market structure

market clearing conditions for the the emission, capacity, and energy markets, and (c) allowances allocation rules. Each of these components is described in detail below.

2.1 Firms' optimization problems

For simplicity, each firm is assumed to own only one type of generating capacity. Taking as exogenous the emission allocation α_f, as well as the the prices (for energy p_t for $t \in \mathcal{T}$, emission allowances p_e, capacity p_c), the firm f solves the following profit maximization problem, whose objective function is revenue less cost, to determine its capacity cap_f and sales s_{ft}:

$$\begin{array}{ll} \underset{\text{cap}_f,\,(s_{ft})_{t\in\mathcal{T}}}{\text{maximize}} & \displaystyle\sum_{t\in\mathcal{T}} H_t\,(\,p_t - \text{MC}_{ft} - p_e\,E_f\,)\,s_{ft} + (\,p_c + p_e\alpha_f - F_f\,)\,\text{cap}_f \\ \text{subject to} & \text{CAP}_f \;\le\; s_{ft} \;\le\; \text{cap}_f, \quad \forall\, t \in \mathcal{T} \end{array} \tag{1}$$

Although the amount of allowances granted per MW of new investment depends on the emissions per MW-year of that capacity type (see the emission rules later), each generator believes (naively) that it cannot affect that amount, and treats it as exogenous. The problem (1) is a linear program whose optimality conditions are straightforward to write down:

$$\begin{array}{lcll} 0 \le \bar{s}_{ft} & \perp & H_t\,(\,-p_t + \text{MC}_{ft} + p_e\,E_f\,) + \mu_{ft} \;\ge\; 0, & \forall\, t \in \mathcal{T} \\ 0 \le \mu_{ft} & \perp & \text{cap}_f - \bar{s}_{ft} - \text{CAP}_f \;\ge\; 0, & \forall\, t \in \mathcal{T} \\ 0 \le \text{cap}_f & \perp & -p_c - p_e\,\alpha_f + F_f - \displaystyle\sum_{t\in\mathcal{T}} \mu_{ft} \;\ge\; 0, & \end{array} \tag{2}$$

where $\perp$ is the perpendicularity notation between two vectors, which in this case simply expresses the complementary slackness condition in linear programming.

When firms instead exert market power, their revenues from energy sales change from linear to nonlinear functions of the sales variables:

$$\sum_{t\in\mathcal{T}} H_t\, s_{ft}\, p_t \;\longrightarrow\; \sum_{t\in\mathcal{T}} H_t\, s_{ft}\, p_t \left(\sum_{g\in\mathcal{F}} s_{gt}\right),$$

and an additional term corresponding to the derivative of the price function $p_t(\bullet)$ with respect to s_{ft} will appear in the first complementarity condition in (2). The rest of the paper focuses on the case where all firms are price-takers. Refinements of the firms' problems are possible; for instance, the model could accommodate spatially distributed generation and sales variables as well as bounded transmission; see, e.g., the previous models [13, 19], as well as linear constraints, such as a "min-run capacity constraint" that is of the form $\mathrm{s}_f \geq \gamma_f \mathrm{cap}_f$ for a firm-dependent constant $\gamma_f > 0$. Nevertheless, the main focus here is on the emission allocation rules to be introduced momentarily, which introduce a new dimension to electric power equilibrium problems that has not been analyzed before. Therefore, we will work with (1) and its equivalent optimality conditions (2) from now on.

2.2 Market clearing conditions

The price p_e of emission allowance is determined by the complementarity condition:

$$0 \;\leq\; p_e \;\perp\; \overline{E} - e_{\mathrm{NP}}(p_e) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, s_{gt} \;\geq\; 0, \tag{3}$$

which stipulates that allowance price is positive only when demand for allowances equals the available supply. Notice that this formulation assumes that allowances can be purchased from or sold to sectors of the economy other than electric power; this is consistent with the EU ETS. The function $e_{\mathrm{NP}}(p_e)$ represents the effective demand for allowances from other sectors [11].

Similarly, the capacity price p_c is determined by the complementarity condition:

$$0 \;\leq\; p_c \;\perp\; \sum_{g\in\mathcal{F}} \mathrm{cap}_g - \underline{\mathrm{CAP}} \;\geq\; 0, \tag{4}$$

which stipulates that capacity price is positive only when demand for capacity equals the supply.

The final market clearing condition stipulates that energy supplies equal the quantity demanded:

$$\sum_{f\in\mathcal{F}} s_{ft} = d_t(p_t), \text{ for all } t \in \mathcal{T}, \text{ or equivalently, } p_t = \pi_t\left(\sum_{f\in\mathcal{F}} s_{ft}\right).$$

Because this condition is an equality, the associated price p_t is unrestricted in sign.

2.3 Emission allocation rules

All emissions rules considered satisfy the condition that the amount of allowances available for allocation equals the amount allocated to capacity:

$$\overline{E} - E_{\mathrm{GF}} \;=\; \sum_{f\in\mathcal{T}} \alpha_f\, \mathrm{cap}_f. \tag{5}$$

We distinguish two types of emission allocation rules for determining α_f, both being based on certain (weighted) averages of CO_2 emission:

(I) The potential emission (or input) rule (in terms of capacity), where

$$\alpha_f \, \mathrm{cap}_f \; = \; \frac{R_f \, \mathrm{cap}_f}{\sum\limits_{g \in \mathcal{F}} R_g \, \mathrm{cap}_g} \, (\, \overline{E} - E_{\mathrm{GF}} \,), \quad \forall \, f \; \in \; \mathcal{F}, \tag{6}$$

provided that the denominator is positive; or equivalently, $\alpha_f = \alpha R_f$ for a common endogenous variable α, due to the allowance allocation balance (5). Under this rule, the emission allowance allocated to new capacity of a particular type is fixed ahead of actual operations and is proportional to the ratio of the firm's capacity to a weighted sum of the capacity owned by all firms.

(II) The actual emission (or output) rule (in terms of sales), where

$$\alpha_f \, \mathrm{cap}_f = \frac{\widehat{R}_f \sum\limits_{t \in \mathcal{T}} H_t \, s_{ft}}{\sum\limits_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_g \, \widehat{R}_g \, s_{gt}} \, (\, \overline{E} - E_{\mathrm{GF}} \,), \quad \forall \, f \; \in \; \mathcal{F}; \tag{7}$$

provided that the denominator is positive; or equivalently, $\alpha_f = \widehat{\alpha} \widehat{R}_f \dfrac{\sum\limits_{t \in \mathcal{T}} H_t s_{ft}}{\mathrm{cap}_f}$ (if $\mathrm{cap}_f > 0$) for a common variable $\widehat{\alpha}$, due to the same emission balancing constraint (5). If the $\widehat{R}_f$ are equal for all f, then this rule allocates allowances in proportion to sales. However, if $\widehat{R}_f$ instead is the emissions rate per MWh, then allowances are allocated in proportion to emissions. In that case, in contrast to rule (I), this rule ensures that if, say, a plant emits 75% of the CO_2, it receives 75% of the allowances $\overline{E} - E_{\mathrm{GF}}$ that are allocated to new capacity.

There are simple conditions ensuring that the denominators in (6) and (7) are positive, such as when $\mathrm{CAP}_f > 0$ for some $f \in \mathcal{F}$. Subsequently, in order for the above rules to be well-defined irrespective of whether the denominator is zero, we write

$$\alpha_f \, \mathrm{cap}_f \; = \; \begin{cases} \alpha \, R_f \, \mathrm{cap}_f & \text{for (6)} \\ \widehat{\alpha} \, \widehat{R}_f \sum\limits_{t \in \mathcal{T}} H_t \, s_{ft} & \text{for (7)} \end{cases} \tag{8}$$

for some nonnegative variables $\widehat{\alpha}$ and α to be determined. Needless to say, other allocation rules are possible, such as some combination of the two expressions in (8), or output-based rules in which allowances are allocated to production rather than to capacity. In the rest of the paper, we focus on rules (I) and (II).

2.4 Model solution

The model seeks a set of electricity sales $(s_{ft})_{(f,t) \in \mathcal{F} \times \mathcal{T}}$, capacities $\left(\mathrm{cap}_f\right)_{f \in \mathcal{F}}$, dual variables $(\mu_{ft})_{(f,t) \in \mathcal{F} \times \mathcal{T}}$, emission allowances $(\alpha_f)_{f \in \mathcal{F}}$, emission allowance price p_e, and capacity price p_c, satisfying, for some nonnegative scalars α and $\widehat{\alpha}$ corresponding the emission allocation rules (I) and (II), respectively, the following

conditions:

$$
\begin{array}{lcll}
0 \le \bar{s}_{ft} & \perp & H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} (\bar{s}_{gt} + \mathrm{CAP}_g) \right) + \mathrm{MC}_{ft} + p_e E_f \right] + \mu_{ft} \ge 0, & \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \le \mu_{ft} & \perp & \mathrm{cap}_f - \bar{s}_{ft} - \mathrm{CAP}_f \ge 0, & \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \le \mathrm{cap}_f & \perp & -p_c - p_e\, \alpha_f + F_f - \sum_{t \in \mathcal{T}} \mu_{ft} \ge 0, & \forall\, f \in \mathcal{F} \\
0 \le p_e & \perp & \overline{E} - e_{\mathrm{NP}}(p_e) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t\, E_g\, (\, \bar{s}_{gt} + \mathrm{CAP}_g\,) \ge 0 & \\
0 \le p_c & \perp & \sum_{g \in \mathcal{F}} \mathrm{cap}_g - \underline{\mathrm{CAP}} \ge 0 & \\
(8) & \text{and} & \sum_{g \in \mathcal{F}} \alpha_g\, \mathrm{cap}_g - (\, \overline{E} - E_{\mathrm{GF}}\,) = 0. &
\end{array}
\tag{9}
$$

3 Nonlinear Complementarity Formulations

To establish the existence of a solution to the model with the emission rules (I) and (II), we derive equivalent formulations of (9) under these rules as standard nonlinear complementarity problems (NCPs).

3.1 The rule (I)

For the emission rule (I): $\alpha_f \mathrm{cap}_f = \alpha R_f \mathrm{cap}_f$ for all $f \in \mathcal{F}$, we introduce a reformulation of (9) that replaces this rule by a complementarity condition. Specifically, we multiply the emission allowance balancing constraint $\sum_{f \in \mathcal{F}} \alpha_f \mathrm{cap}_f = \overline{E} - E_{\mathrm{GF}}$ by the variable p_e, turn the resulting equation into an inequality, introduce the nonnegative variable $\sigma \equiv \alpha p_e$, and impose complementarity between σ and the modified emission inequality. The resulting NCP is as follows:

$$
\begin{array}{lcll}
0 \le \bar{s}_{ft} & \perp & H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} (\bar{s}_{gt} + \mathrm{CAP}_g) \right) + \mathrm{MC}_{ft} + p_e E_f \right] + \mu_{ft} \ge 0, & \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \le \mu_{ft} & \perp & \mathrm{cap}_f - \bar{s}_{ft} - \mathrm{CAP}_f \ge 0, & \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \le \mathrm{cap}_f & \perp & -p_c - \sigma\, R_f + F_f - \sum_{t \in \mathcal{T}} \mu_{ft} \ge 0, & \forall\, f \in \mathcal{F} \\
0 \le p_e & \perp & \overline{E} - e_{\mathrm{NP}}(p_e) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t\, E_g\, (\, \bar{s}_{gt} + \mathrm{CAP}_g\,) \ge 0 & \\
0 \le p_c & \perp & \sum_{g \in \mathcal{F}} \mathrm{cap}_g - \underline{\mathrm{CAP}} \ge 0 & \\
0 \le \sigma & \perp & \sigma \sum_{g \in \mathcal{F}} R_g\, \mathrm{cap}_g - (\, \overline{E} - E_{\mathrm{GF}}\,)\, p_e \ge 0. &
\end{array}
\tag{10}
$$

Note the difference between (9) and (10): the former is a mixed NCP in the variable $(\mathbf{x}, \alpha)$, where

$$
\mathbf{x} \equiv \left\{ (s_{ft})_{(f,t) \in \mathcal{F} \times \mathcal{T}}, (\mu_{ft})_{(f,t) \in \mathcal{F} \times \mathcal{T}}, \left(\mathrm{cap}_f\right)_{f \in \mathcal{F}}, (\alpha_f)_{f \in \mathcal{F}}, p_e, p_c \right\};
$$

the latter is a standard NCP in the variable

$$
\mathbf{x}^{\mathrm{I}} \equiv \left\{ (s_{ft})_{(f,t) \in \mathcal{F} \times \mathcal{T}}, (\mu_{ft})_{(f,t) \in \mathcal{F} \times \mathcal{T}}, \left(\mathrm{cap}_f\right)_{f \in \mathcal{F}}, p_e, p_c, \sigma \right\}.
$$

The following result summarizes the connection between (10) and (9) with the emission rule (I) under the mild condition (11), which postulates that, in the event where $\underline{\mathrm{CAP}} + \sum_{f\in\mathcal{F}} \mathrm{CAP}_f = 0$, at least one firm's investment cost for capacity is not too high.

Proposition 1. Assume that

$$\max\left(\underline{\mathrm{CAP}} + \sum_{f\in\mathcal{F}} \mathrm{CAP}_f,\ \chi \max_{f\in\mathcal{F}} \left\{ \sum_{t\in\mathcal{T}} H_t\,[\,\pi_t(0) - \mathrm{MC}_{ft}\,] - F_f \right\}\right) > 0. \tag{11}$$

If $(\mathbf{x}, \alpha)$ is a solution of (9) under the emission rule (I), then

$$\sigma \equiv \frac{(\,\overline{E} - E_{\mathrm{GF}}\,)\,p_e}{\sum_{g\in\mathcal{F}} R_g\,\mathrm{cap}_g} \tag{12}$$

is well defined and $\mathbf{x}^{\mathrm{I}}$ is a solution of (10). Conversely, if $\mathbf{x}^{\mathrm{I}}$ is a solution of (10), then

$$\alpha \equiv \frac{\overline{E} - E_{\mathrm{GF}}}{\sum_{g\in\mathcal{F}} R_g\,\mathrm{cap}_g} \tag{13}$$

is well defined, and with $\alpha_f \equiv \alpha R_f$ for all $f \in \mathcal{F}$, $(\mathbf{x}, \alpha)$ is a solution of (9) under the emission rule (I).

Proof. To prove the first assertion, let $(\mathbf{x}, \alpha)$ be as given. We first show that $\mathrm{cap}_f > 0$ for some $f \in \mathcal{F}$. Suppose not, then $\mathrm{cap}_f = s_{ft} = 0$ for all $(f, t) \in \mathcal{F} \times \mathcal{T}$, which implies $\underline{\mathrm{CAP}} + \sum_{f\in\mathcal{F}} \mathrm{CAP}_f = 0$. From the first and third line in (9), we deduce, for all $f \in \mathcal{F}$,

$$\sum_{t\in\mathcal{T}} H_t\,[\,-\pi_t(0) + \mathrm{MC}_{ft}\,] + F_f \ \geq\ 0;$$

or equivalently,

$$\max_{f\in\mathcal{F}} \left\{ \sum_{t\in\mathcal{T}} H_t\,[\,\pi_t(0) - \mathrm{MC}_{ft}\,] - F_f \right\} \leq\ 0,$$

which contradicts (11). Therefore, the scalar σ in (12) is well defined; moreover, $\sigma = p_e\alpha$, yielding $\sigma R_f = p_e\alpha_f$. Consequently, (10) follows from (9). Conversely, let $\mathbf{x}^{\mathrm{I}}$ be a solution of (10). By the same argument as before, we deduce that $\mathrm{cap}_f > 0$ for some $f \in \mathcal{F}$. Therefore, the scalar α in (13) is well defined; let $\alpha_f \equiv \alpha R_f$. We then have $\sum_{f\in\mathcal{F}} \alpha_f \mathrm{cap}_f = \overline{E} - E_{\mathrm{GF}}$. Consequently, $(\mathbf{x}, \alpha)$ is a solution (9) under the emission rule (I). □

Condition (11) implies that if $\underline{\mathrm{CAP}} + \sum_{f\in\mathcal{F}} \mathrm{CAP}_f = 0$, then $\max_{f\in\mathcal{F}} \left\{ \sum_{t\in\mathcal{T}} H_t\,[\,\pi_t(0) - \mathrm{MC}_{ft}\,] - F_f \right\} > 0$, which allows each firm to produce zero power. If no firm sells any power, then the power price will be expected to very high at each time interval. Therefore, it is reasonable to assume that there is at least one firm which will find it profitable to invest; i.e., whose total short-run and investment cost is less than their revenue, on a per MW of investment basis.

It turns out that if $\underline{\mathrm{CAP}} + \sum_{f\in\mathcal{F}} \mathrm{CAP}_f > 0$, then the NCP (10) is equivalent to the set of Karush-Kuhn-Tucker (KKT) conditions of the variational inequality (VI) defined by the pair $(K^{\mathrm{I}}, \Phi^{\mathrm{I}})$, where $K^{\mathrm{I}} \equiv K \times \Re_+$ with

$$K \equiv \left\{ (\,\bar{\mathbf{s}}, \mathbf{cap}\,) \,\geq\, 0 \,:\, \sum_{g\in\mathcal{F}} \mathrm{cap}_g - \underline{\mathrm{CAP}} \,\geq\, 0 \;\text{ and }\; \mathrm{cap}_f - \bar{s}_{ft} - \mathrm{CAP}_f \,\geq\, 0,\; \forall\,(f,t) \in \mathcal{F}\times\mathcal{T} \right\}$$

being an unbounded polyhedron in the variables $\bar{\mathbf{s}} \equiv (\bar{s}_{ft})_{(f,t)\in\mathcal{F}\times\mathcal{T}}$ and $\mathbf{cap} \equiv \left(\mathrm{cap}_f\right)_{f\in\mathcal{F}}$, and

$$\Phi^{\mathrm{I}}(\bar{\mathbf{s}}, \mathbf{cap}, p_e) \,\equiv\, \left(\begin{array}{c} \left(H_t \left[-\pi_t \left(\sum_{g\in\mathcal{F}} (\,\bar{s}_{gt} + \mathrm{CAP}_g\,) \right) + \mathrm{MC}_f + p_e\, E_f \right] \right)_{(f,t)\in\mathcal{F}\times\mathcal{T}} \\ \left(F_f - \dfrac{(\,\overline{E} - E_{\mathrm{G}F}\,)\, p_e}{\sum_{g\in\mathcal{F}} R_g\, \mathrm{cap}_g} R_f \right)_{f\in\mathcal{F}} \\ \overline{E} - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, (\,\bar{s}_{gt} + \mathrm{CAP}_g\,) \,-\, e_{\mathrm{N}P}(p_e) \end{array} \right)$$

is a non-monotone map. Indeed, note that Φ^{I} is well defined on the set K^{I} because for every element $(\bar{\mathbf{s}}, \mathbf{cap}) \in K$, we must have $\mathrm{cap}_f > 0$ for some $f \in \mathcal{F}$. Letting p_c and μ_{ft} be the multipliers of the functional constraints in K, we can readily write down the KKT conditions of the VI $(K^{\mathrm{I}}, \Phi^{\mathrm{I}})$ and conclude that they are equivalent to the NCP (10) under the identification (12) for σ. When $\mathrm{CAP} = 0 < \sum_{f\in\mathcal{F}} \mathrm{CAP}_f$, the set

$$K \,=\, \prod_{f\in\mathcal{F}} \left\{ \left((\,s_{ft}\,)_{t\in\mathcal{T}}, \mathrm{cap}_f \right) \,\geq\, 0 \,:\, \mathrm{cap}_f - \bar{s}_{ft} - \mathrm{CAP}_f \,\geq\, 0,\; \forall\, t \in \mathcal{T} \right\}$$

is the Cartesian product of separable sets. While the VI formulation is quite compact, one obvious advantage of the NCP (10) is that it applies to the case where $\underline{\mathrm{CAP}} = \mathrm{CAP}_f = 0$ for all $f \in \mathcal{F}$; in the latter case, the set K^{I} contains the origin where the function Φ^{I} fails to be well defined.

3.2 The emission rule (II)

The NCP formulation for (9) under the emission rule (II) is somewhat different. For one thing, there is no change of variables in the formulation; in particular, the variables α_f are kept in the formulation.

The derivation in this subsection permits $\underline{\text{CAP}} + \sum_{f\in\mathcal{F}} \text{CAP}_f = 0$. Specifically, consider the NCP:

$$
\begin{array}{lcll}
0 \le \bar{s}_{ft} & \perp & H_t \left[-\pi_t \left(\sum_{g\in\mathcal{F}} (\bar{s}_{gt} + \text{CAP}_g) \right) + \text{MC}_{ft} + p_e E_f \right] + \mu_{ft} \ge 0, & \forall\, (f,t) \in \mathcal{F}\times\mathcal{T} \\
0 \le \mu_{ft} & \perp & \text{cap}_f - \bar{s}_{ft} - \text{CAP}_f \ge 0, & \forall\, (f,t) \in \mathcal{F}\times\mathcal{T} \\
0 \le \text{cap}_f & \perp & -p_c - \alpha_f\, p_e + F_f - \sum_{t\in\mathcal{T}} \mu_{ft} \ge 0, & \forall\, f \in \mathcal{F} \\
0 \le \alpha_f & \perp & \alpha_f\, \text{cap}_f - \widehat{\alpha}\, \widehat{R}_f \sum_{t\in\mathcal{T}} H_t\, (\, \bar{s}_{ft} + \text{CAP}_f\,) \ge 0, & \forall\, f \in \mathcal{F} \\
0 \le p_e & \perp & \overline{E} - e_{NP}(p_e) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, (\, \bar{s}_{gt} + \text{CAP}_g\,) \ge 0 & \\
0 \le p_c & \perp & \sum_{g\in\mathcal{F}} \text{cap}_g - \underline{\text{CAP}} \ge 0 & \\
0 \le \widehat{\alpha} & \perp & \sum_{g\in\mathcal{F}} \alpha_g\, \text{cap}_g - (\, \overline{E} - E_{\text{GF}}\,) \ge 0 &
\end{array}
\tag{14}
$$

in the variable $(\mathbf{x}, \widehat{\alpha})$. The following result establishes the equivalence of the above NCP with (9) under the emission rule (II).

Proposition 2. A pair $(\mathbf{x}, \widehat{\alpha})$ is a solution of (9) under the emission rule (II) if and only if it is a solution of (14).

Proof. The "only if" statement is obvious. Conversely, if $(\mathbf{x}, \alpha)$ is a solution of (14), it suffices to show that the following equalities hold:

$$
\sum_{g\in\mathcal{F}} \alpha_g\, \text{cap}_g - (\, \overline{E} - E_{\text{GF}}\,) = 0 \quad \text{and} \quad \alpha_f\, \text{cap}_f - \widehat{\alpha}\, \widehat{R}_f \sum_{t\in\mathcal{T}} H_t\, E_f\, (\, \bar{s}_{ft} + \text{CAP}_f\,) = 0,\ \forall\, f \in \mathcal{F}.
$$

Indeed if the first equality does not hold, then $\widehat{\alpha} = 0$ by complementarity, yielding $0 \le \alpha_f \perp \alpha_f \text{cap}_f \ge 0$. In turn, this implies $\alpha_f \text{cap}_f = 0$ for all $f \in \mathcal{F}$; thus $\overline{E} - E_{\text{GF}} = 0$, which contradicts the assumption that $\overline{E} > E_{\text{GF}}$. Similarly, if $\alpha_f \text{cap}_f - \widehat{\alpha}\widehat{R}_f \sum_{t\in\mathcal{T}} H_t E_f (\bar{s}_{ft} + \text{CAP}_f) > 0$ for some $f \in \mathcal{F}$, then $\alpha_f = 0$ by complementarity, which contradicts the inequality itself. □

Similar to the VI $(K^{\text{I}}, \Phi^{\text{I}})$, if $\text{CAP}_f > 0$ for all $f \in \mathcal{F}$, the NCP (14) is equivalent to the KKT conditions of the VI $(K^{\text{II}}, \Phi^{\text{II}})$, where $K^{\text{II}} \equiv K^{\text{I}} = K \times \Re_+$ and

$$
\Phi^{\text{II}}(\bar{\mathbf{s}}, \mathbf{cap}, p_e) \equiv \left(\begin{array}{c}
\left(H_t \left[-\pi_t \left(\sum_{g\in\mathcal{F}} (\, \bar{s}_{gt} + \text{CAP}_g\,) \right) + \text{MC}_f + p_e\, E_f \right] \right)_{(f,t)\in\mathcal{F}\times\mathcal{T}} \\
\left(F_f - \dfrac{(\, \overline{E} - E_{\text{GF}}\,)\, p_e\, \widehat{R}_f \sum_{t\in\mathcal{T}} H_t\, (\, \bar{s}_{ft} + \text{CAP}_f\,)}{\text{cap}_f \sum_{g\in\mathcal{F}} \widehat{R}_g \sum_{t\in\mathcal{T}} H_t\, (\bar{s}_{gt} + \text{CAP}_g\,)} \right)_{f\in\mathcal{F}} \\
\overline{E} - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, (\, \bar{s}_{gt} + \text{CAP}_g\,) - e_{NP}(p_e)
\end{array} \right).
$$

4 Existence of Solutions

For the analysis in this section, we introduce the following mild condition on the supply of allowances:

$$\overline{E} \;>\; e_{\mathrm{NP}}(0) + \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, \mathrm{CAP}_g. \tag{15}$$

This condition merely says that the total number of allowances is sufficient to cover the emissions resulting from the sum across firms of the lower bounds upon generation, net of the supply of allowances from other sectors at an allowance price of zero. This is not restrictive, because at a price of zero, it is likely that the supply from other sectors is nonpositive (as other sectors would likely be willing to buy allowances at such a low price), and because the minimum required generation is likely to be a small fraction of total generation, thereby requiring few allowances.

We also impose the following condition that is a slight strengthening of (11):

$$\max_{f\in\mathcal{F}} \left\{ \sum_{t\in\mathcal{T}} H_t \left[\pi_t \left(\sum_{g\in\mathcal{F}} \mathrm{CAP}_g \right) - \mathrm{MC}_{ft} \right] - F_f \right\} \;>\; 0. \tag{16}$$

This condition states that at least one firm would find investment profitable (fixed and variable cost less than revenue) if every generator is producing just its individual lower bound. This is a mild restriction, as the power price would likely be very high when everyone is producing at their lowest possible level.

The following proposition shows that under these two conditions, any solution to the model (9) is nontrivial.

Proposition 3. Under (15) and (16), any solution of the NCP (10) and (14) must have $\bar{s}_{ft} > 0$ for some $(f,t) \in \mathcal{F}\times\mathcal{T}$.

Proof. We prove the proposition only for (14). Assume for the sake of contradiction that some solution of this NCP has $\bar{s}_{ft} = 0$ for all $(f,t) \in \mathcal{F}\times\mathcal{T}$. We claim that $p_e = 0$. Indeed, if $p_e > 0$, then

$$0 \;=\; \overline{E} - e_{\mathrm{NP}}(p_e) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, \mathrm{CAP}_g \;\geq\; \overline{E} - e_{\mathrm{NP}}(0) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, \mathrm{CAP}_g \;>\; 0,$$

which is a contradiction. Hence, we have, for each $f \in \mathcal{T}$,

$$\begin{aligned} 0 \;&\leq\; \sum_{t\in\mathcal{T}} \left\{ H_t \left[-\pi_t \left(\sum_{g\in\mathcal{F}} \mathrm{CAP}_g \right) + \mathrm{MC}_{ft} \right] + \mu_{ft} \right\} \\ &\leq\; \sum_{t\in\mathcal{T}} H_t \left[-\pi_t \left(\sum_{g\in\mathcal{F}} \mathrm{CAP}_g \right) + \mathrm{MC}_{ft} \right] + F_f, \end{aligned}$$

which contradicts (16). □

We recall that the temporal price functions $\pi_t(\bullet)$ are strictly decreasing and the nonpower emission function $e_{\mathrm{NP}}(\bullet)$ is nonincreasing. The following is the main existence theorem for the model (9) with the emission rules (I) and (II).

Theorem 4. Under conditions (15) and (16), the model (9) has a solution.

We prove the above theorem via the two equivalent NCPs: (10) for emission rule (I) and (14) for emission rule (II). In turn, the proofs for these two NCPs are quite similar. Both are based on the application of a fundamental existence result for a general NCP summarized below. A proof of this lemma can be found in [6, Theorem 2.6.1].

Lemma 5. Let $\Phi : \Re^n \to \Re^n$ be a continuous function. If there exists a constant $c > 0$ such that all solutions of the NCP: $0 \leq x \perp \Phi(x) + \tau x \geq 0$ for $\tau > 0$ satisfy $\|x\| \leq c$, then the NCP: $0 \leq x \perp \Phi(x) \geq 0$ has a solution. □

To avoid repetition, we present the proof for the NCP (14) only; see Subsection 4.1. We choose this NCP because there is an extra perturbation step that is needed in applying the lemma, whereas one can follow the same argument and directly apply the lemma to the NCP (10).

4.1 Proof for the NCP (14)

Toward the proof of solution existence to the NCP (14), we consider a perturbation of the function $\Phi^{\rm II}$ in order to deal with the general case where some $\mathrm{CAP}_f = 0$. Specifically, for each $\varepsilon > 0$, let

$$\Phi^{\rm II}_\varepsilon(\bar{\mathbf{s}}, \mathbf{cap}, p_e) \equiv \begin{pmatrix} \left(H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} (\bar{s}_{gt} + \mathrm{CAP}_g) \right) + \mathrm{MC}_f + p_e E_f \right] \right)_{(f,t) \in \mathcal{F} \times \mathcal{T}} \\ \left(F_f - \dfrac{(\overline{E} - E_{\rm GF}) p_e \widehat{R}_f \sum_{t \in \mathcal{T}} H_t (\bar{s}_{ft} + \mathrm{CAP}_f)}{(\mathrm{cap}_f + \varepsilon) \left(\varepsilon + \sum_{g \in \mathcal{F}} \widehat{R}_g \sum_{t \in \mathcal{T}} H_t (\bar{s}_{gt} + \mathrm{CAP}_g) \right)} \right)_{f \in \mathcal{F}} \\ \overline{E} - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t E_g (\bar{s}_{gt} + \mathrm{CAP}_g) - e_{\rm NP}(p_e) \end{pmatrix},$$

which is well defined on the set $K^{\rm II}$. We first show that the VI $(K^{\rm II}, \Phi^{\rm II}_\varepsilon)$ has a solution for each fixed but arbitrary $\varepsilon > 0$. For this purpose, we take an arbitrary sequence of positive scalars $\{\tau_k\}$; for each k, let $(\bar{\mathbf{s}}^{\varepsilon,k}, \mathbf{cap}^{\varepsilon,k}, \boldsymbol{\mu}^{\varepsilon,k}, p_e^{\varepsilon,k}, p_c^{\varepsilon,k})$ be a tuple satisfying

$$0 \leq \bar{s}^{\varepsilon,k}_{ft} \quad \perp \; H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} (\bar{s}^{\varepsilon,k}_{gt} + \mathrm{CAP}_g) \right) + \mathrm{MC}_{ft} + p_e^{\varepsilon,k} E_f \right] + \mu^{\varepsilon,k}_{ft} + \tau_k \bar{s}^{\varepsilon,k}_{ft} \geq 0, \; \forall\, (f,t) \in \mathcal{F} \times \mathcal{T}$$

$$0 \leq \mu^{\varepsilon,k}_{ft} \quad \perp \; \mathrm{cap}^{\varepsilon,k}_f - \bar{s}^{\varepsilon,k}_{ft} - \mathrm{CAP}_f + \tau_k \mu^{\varepsilon,k}_{ft} \; \geq \; 0, \quad \forall\, (f,t) \in \mathcal{F} \times \mathcal{T}$$

$$0 \leq \mathrm{cap}^{\varepsilon,k}_f \; \perp \; -p_c^{\varepsilon,k} + F_f - \frac{(\overline{E} - E_{\rm GF}) p_e^{\varepsilon,k} \widehat{R}_f \sum_{t \in \mathcal{T}} H_t (\bar{s}^{\varepsilon,k}_{ft} + \mathrm{CAP}_f)}{(\mathrm{cap}^{\varepsilon,k}_f + \varepsilon) \left(\varepsilon + \sum_{g \in \mathcal{F}} \widehat{R}_g \sum_{t \in \mathcal{T}} H_t (\bar{s}^{\varepsilon,k}_{gt} + \mathrm{CAP}_g) \right)} - \sum_{t \in \mathcal{T}} \mu^{\varepsilon,k}_{ft} + \tau_k \mathrm{cap}^{\varepsilon,k}_f \geq 0,$$
$$\forall\, f \in \mathcal{F}$$

$$0 \leq p_e^{\varepsilon,k} \quad \perp \; \overline{E} - e_{\rm NP}(p_e^{\varepsilon,k}) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t E_g \left(\bar{s}^{\varepsilon,k}_{gt} + \mathrm{CAP}_g \right) + \tau_k p_e^{\varepsilon,k} \; \geq \; 0$$

$$0 \leq p_c^{\varepsilon,k} \quad \perp \; \sum_{g \in \mathcal{F}} \mathrm{cap}^{\varepsilon,k}_g - \underline{\mathrm{CAP}} + \tau_k p_c^{\varepsilon,k} \; \geq \; 0.$$

We claim that under condition (15), the sequence $\{(\bar{\mathbf{s}}^{\varepsilon,k}, \mathbf{cap}^{\varepsilon,k}, \boldsymbol{\mu}^{\varepsilon,k}, p_e^{\varepsilon,k}, p_c^{\varepsilon,k})\}$ is bounded. We show this in several steps: first the sequence $\{p_e^{\varepsilon,k}\}$; next the sequence $\{\bar{s}_{ft}^{\varepsilon,k}\}$ for all $(f,t) \in \mathcal{F} \times \mathcal{T}$; then the sequence $\{\mathrm{cap}_f^{\varepsilon,k}\}$ for all $f \in \mathcal{F}$.

Boundedness of $\{p_e^{\varepsilon,k}\}$. Assume for the sake of contradiction that $\{p_e^{\varepsilon,k}\}$ is unbounded. Then for an infinite index set $\kappa \subset \{1, 2, \cdots, \infty\}$, we have

$$\lim_{k(\in\kappa)\to\infty} p_e^{\varepsilon,k} \;=\; \infty. \tag{17}$$

Without loss of generality, we may assume that $p_e^{\varepsilon,k} > 0$ for all $k \in \kappa$. It follows by complementarity that

$$\overline{E} - e_{\mathrm{NP}}(p_e^{\varepsilon,k}) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g \left(\bar{s}_{gt}^{\varepsilon,k} + \mathrm{CAP}_g \right) + \tau_k\, p_e^{\varepsilon,k} \;=\; 0, \quad \forall\, k \,\in\, \kappa.$$

For any $k \in \kappa$ such that $\bar{s}_{f_0t_0}^{\varepsilon,k} > 0$ for some pair (f_0, t_0), we have

$$H_{t_0} \left[-\pi_{t_0} \left(\sum_{g\in\mathcal{F}} (\bar{s}_{gt_0}^{\varepsilon,k} + \mathrm{CAP}_g) \right) + \mathrm{MC}_{f_0t_0} + p_e^{\varepsilon,k}\, E_{f_0} \right] + \mu_{f_0t_0}^{\varepsilon,k} + \tau_k \bar{s}_{f_0t_0}^{\varepsilon,k} \;=\; 0,$$

which yields

$$p_e^{\varepsilon,k} \;\le\; E_{f_0}^{-1} \left[\pi_{t_0} \left(\sum_{g\in\mathcal{F}} \mathrm{CAP}_g \right) - \mathrm{MC}_{f_0t_0} \right]. \tag{18}$$

On the other hand, if $k \in \kappa$ is such that $\bar{s}_{ft}^{\varepsilon,k} = 0$ for all (f,t), then we have

$$\begin{aligned} 0 \;&=\; \overline{E} - e_{\mathrm{NP}}(p_e^{\varepsilon,k}) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, \mathrm{CAP}_g + \tau_k\, p_e^{\varepsilon,k} \\ &\ge\; \overline{E} - e_{\mathrm{NP}}(0) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, \mathrm{CAP}_g \;>\; 0, \end{aligned}$$

which contradicts (15). Consequently, the bound (18) holds for all $k \in \kappa$; contradicting the limit (17).

Boundedness of $\{\bar{s}_{ft}^{\varepsilon,k}\}$ **for every** $(f,t) \in \mathcal{F} \times \mathcal{T}$**.** Assume for the sake of contradiction that for some pair (f_0, t_0) and an infinite set $\kappa \subset \{1, 2, \cdots, \}$,

$$\lim_{k(\in\kappa)\to\infty} \bar{s}_{f_0t_0}^{\varepsilon,k} \;=\; \infty. \tag{19}$$

Without loss of generality, we may assume that $\bar{s}_{f_0t_0}^{\varepsilon,k} > 0$ for all $k \in \kappa$. It follows by complementarity that

$$\begin{aligned} 0 \;&=\; H_{t_0} \left[-\pi_{t_0} \left(\sum_{g\in\mathcal{F}} (\bar{s}_{gt_0}^{\varepsilon,k} + \mathrm{CAP}_g) \right) + \mathrm{MC}_{f_0t_0} + p_e^{\varepsilon,k}\, E_{f_0} \right] + \mu_{f_0t_0}^{\varepsilon,k} + \tau_k\, \bar{s}_{f_0t_0}^{\varepsilon,k} \\ &\ge\; H_{t_0} \left[-\pi_{t_0} \left(\sum_{g\in\mathcal{F}} \mathrm{CAP}_g \right) + \mathrm{MC}_{ft} + p_e^{\varepsilon,k}\, E_{f_0} \right] + \max(\, \mu_{f_0t_0}^{\varepsilon,k},\, \tau_k\, \bar{s}_{f_0t_0}^{\varepsilon,k} \,) \end{aligned}$$

which implies

$$\max(\, \mu_{f_0t_0}^{\varepsilon,k},\, \tau_k\, \bar{s}_{f_0t_0}^{\varepsilon,k} \,) \;\le\; H_{t_0} \left[\pi_{t_0} \left(\sum_{g\in\mathcal{F}} \mathrm{CAP}_g \right) - \mathrm{MC}_{ft} - p_e^{\varepsilon,k}\, E_{f_0} \right].$$

Since the right-hand side is bounded, by (19), it follows that

$$\lim_{k(\in\kappa)\to\infty} \tau_k \;=\; 0. \tag{20}$$

But this contradicts the inequality: $\overline{E} - e_{\rm NP}(p_e^{\varepsilon,k}) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t E_g \left(\bar{s}_{gt}^{\varepsilon,k} + {\rm CAP}_g\right) + \tau_k p_e^{\varepsilon,k} \geq 0$. Therefore, $\{\bar{s}_{ft}^{\varepsilon,k}\}$ is bounded for all $(f,t) \in \mathcal{F}\times\mathcal{T}$.

Boundedness of $\{{\rm cap}_f^{\varepsilon,k}\}$ for every $f \in \mathcal{F}$. Assume for the sake of contradiction that for some f_0 and an infinite set $\kappa \subset \{1, 2, \cdots, \}$,

$$\lim_{k(\in\kappa)\to\infty} {\rm cap}_{f_0}^{\varepsilon,k} \;=\; \infty. \tag{21}$$

Thus, by complementarity, we must have $p_c^{\varepsilon,k} = 0$ for all $k \in \kappa$ sufficiently large. Since $\{\bar{s}_{f_0t}^{\varepsilon,k}\}$ is bounded, we deduce $\mu_{f_0t}^{\varepsilon,k} = 0$ for all $t \in \mathcal{T}$ and all $k \in \kappa$ sufficiently large. Without loss of generality, we may assume that ${\rm cap}_{f_0}^{\varepsilon,k} > 0$ for all $k \in \kappa$. It follows by complementarity that for all $k \in \kappa$ sufficiently large,

$$F_{f_0} - \frac{(\overline{E} - E_{\rm GF})\, p_e^{\varepsilon,k}\, \widehat{R}_{f_0} \sum_{t\in\mathcal{T}} H_t\, (\, \bar{s}_{f_0t}^{\varepsilon,k} + {\rm CAP}_{f_0}\,)}{(\, {\rm cap}_{f_0}^{\varepsilon,k} + \varepsilon\,) \left(\, \varepsilon + \sum_{g\in\mathcal{F}} \widehat{R}_g \sum_{t\in\mathcal{T}} H_t(\, \bar{s}_{gt}^{\varepsilon,k} + {\rm CAP}_g)\right)} + \tau_k\, {\rm cap}_{f_0}^{\varepsilon,k} \;=\; 0,$$

which implies that

$$F_{f_0} \;\leq\; \frac{(\overline{E} - E_{\rm GF})\, p_e^{\varepsilon,k}\, \widehat{R}_{f_0} \sum_{t\in\mathcal{T}} H_t\, (\, \bar{s}_{f_0t}^{\varepsilon,k} + {\rm CAP}_{f_0}\,)}{(\, {\rm cap}_{f_0}^{\varepsilon,k} + \varepsilon\,) \left(\, \varepsilon + \sum_{g\in\mathcal{F}} \widehat{R}_g \sum_{t\in\mathcal{T}} H_t(\, \bar{s}_{gt}^{\varepsilon,k} + {\rm CAP}_g)\right)}.$$

The limit (21) implies that the right-hand side tends to zero as $k(\in\kappa) \to \infty$, which is a contradiction. Hence $\{{\rm cap}_f^{\varepsilon,k}\}$ is bounded for all $f \in \mathcal{F}$.

Boundedness of $\{\mu_{ft}^{\varepsilon,k}\}$ for all $(f,t) \in \mathcal{F}\times\mathcal{T}$. Assume for the sake of contradiction that for some $(f_0,t_0) \in \mathcal{F}\times\mathcal{T}$ and an infinite set $\kappa \subset \{1, 2, \cdots, \}$,

$$\lim_{k(\in\kappa)\to\infty} \mu_{f_0t_0}^{\varepsilon,k} \;=\; \infty. \tag{22}$$

Without loss of generality we may assume that $\mu_{f_0t_0}^{\varepsilon,k} > 0$ for all $k \in \kappa$. By complementarity, we deduce ${\rm cap}_{f_0}^{\varepsilon,k} - \bar{s}_{f_0t_0}^{\varepsilon,k} + {\rm CAP}_{f_0} + \tau_k \mu_{f_0t_0}^{\varepsilon,k} = 0$. Since $\{({\rm cap}_{f_0}^{\varepsilon,k}, \bar{s}_{f_0t_0}^{\varepsilon,k})\}$ is bounded, (22) implies that

$$\lim_{k(\in\kappa)\to\infty} \tau_k \;=\; 0.$$

Since

$$\begin{aligned}
\mu_{f_0t_0}^{\varepsilon,k} \leq \sum_{t\in\mathcal{T}} \mu_{f_0t}^{\varepsilon,k} \;&\leq\; -p_c^{\varepsilon,k} + F_{f_0} - \frac{(\overline{E} - E_{\rm GF}) p_e^{\varepsilon,k}\, \widehat{R}_{f_0} \sum_{t\in\mathcal{T}} H_t(\bar{s}_{f_0t}^{\varepsilon,k} + {\rm CAP}_{f_0})}{({\rm cap}_{f_0}^{\varepsilon,k} + \varepsilon) \left(\, \varepsilon + \sum_{g\in\mathcal{F}} \widehat{R}_g \sum_{t\in\mathcal{T}} H_t(\bar{s}_{gt}^{\varepsilon,k} + {\rm CAP}_g)\right)} + \tau_k {\rm cap}_{f_0}^{\varepsilon,k} \\
&\leq\; F_{f_0} + \tau_k\, {\rm cap}_{f_0}^{\varepsilon,k}
\end{aligned}$$

and $\tau_k \mathrm{cap}_{f_0}^{\varepsilon,k} \to 0$ as $k(\in \kappa) \to \infty$, we obtain a contradiction to (22).

Boundedness of $\{p_c^{\varepsilon,k}\}$. This is similar to the above proof of the μ-sequence.

We have therefore completed the proof of the boundedness of the sequence $\{(\bar{\mathbf{s}}^{\varepsilon,k}, \mathbf{cap}^{\varepsilon,k}, \boldsymbol{\mu}^{\varepsilon,k}, p_e^{\varepsilon,k}, p_c^{\varepsilon,k})\}$ under the condition (15). This is enough to apply Lemma 5 to deduce the existence of a solution to the VI $(K^{\mathrm{II}}, \Phi_\varepsilon^{\mathrm{II}})$ for all $\varepsilon > 0$ via its KKT formulation. Let $(\bar{\mathbf{s}}^\varepsilon, \mathbf{cap}^\varepsilon, p_e^\varepsilon)$ be one such solution. For each $\varepsilon > 0$, there exists $(\mu_{ft}^\varepsilon, p_c^\varepsilon)$ such that

$$
\begin{aligned}
0 \le \bar{s}_{ft}^\varepsilon \quad &\perp \quad H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} (\bar{s}_{gt}^\varepsilon + \mathrm{CAP}_g) \right) + \mathrm{MC}_{ft} + p_e^\varepsilon E_f \right] + \mu_{ft}^\varepsilon \ \ge \ 0, \ \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \le \mu_{ft}^\varepsilon \quad &\perp \quad \mathrm{cap}_f^\varepsilon - \bar{s}_{ft}^\varepsilon - \mathrm{CAP}_f \ \ge \ 0, \quad \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \le \mathrm{cap}_f^\varepsilon \quad &\perp \quad -p_c^\varepsilon + F_f - \frac{(\overline{E} - E_{\mathrm{GF}})\, p_e^\varepsilon\, \widehat{R}_f \sum_{t \in \mathcal{T}} H_t\, (\, \bar{s}_{ft}^\varepsilon + \mathrm{CAP}_f\,)}{(\,\mathrm{cap}_f^\varepsilon + \varepsilon\,) \left(\varepsilon + \sum_{g \in \mathcal{F}} \widehat{R}_g \sum_{t \in \mathcal{T}} H_t\, (\, \bar{s}_{gt}^\varepsilon + \mathrm{CAP}_g\,) \right)} - \sum_{t \in \mathcal{T}} \mu_{ft}^\varepsilon \ge 0, \quad \forall\, f \in \mathcal{F} \\
0 \le p_e^\varepsilon \quad &\perp \quad \overline{E} - e_{\mathrm{NP}}(p_e^\varepsilon) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t\, E_g\, \left(\bar{s}_{gt}^\varepsilon + \mathrm{CAP}_g \right) \ \ge \ 0 \\
0 \le p_c^\varepsilon \quad &\perp \quad \sum_{g \in \mathcal{F}} \mathrm{cap}_g^\varepsilon - \underline{\mathrm{CAP}} \ \ge \ 0.
\end{aligned}
$$

By the same proof sequence as before, we can show that $\limsup_{\varepsilon \downarrow 0} \| (\, \bar{\mathbf{s}}^\varepsilon, \mathbf{cap}^\varepsilon, \boldsymbol{\mu}^\varepsilon, p_e^\varepsilon, p_c^\varepsilon\,) \| \ < \ \infty$. Let $(\widehat{\mathbf{s}}, \widehat{\mathbf{cap}}, \widehat{\boldsymbol{\mu}}, \widehat{p}_e, \widehat{p}_c)$ be the limit of a convergence sequence $\{(\bar{\mathbf{s}}^{\varepsilon_k}, \mathbf{cap}^{\varepsilon_k}, \boldsymbol{\mu}^{\varepsilon_k}, p_e^{\varepsilon_k}, p_c^{\varepsilon_k})\}$ corresponding to a sequence of positive scalars $\{\varepsilon_k\} \downarrow 0$. It follows ready that $(\widehat{\bar{\mathbf{s}}}, \widehat{\mathbf{cap}}, \widehat{\boldsymbol{\mu}}, \widehat{p}_e, \widehat{p}_c)$ satisfies

$$
\begin{aligned}
0 \ \le \ \widehat{\bar{s}}_{ft} \quad &\perp \quad H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} (\, \widehat{\bar{s}}_{gt} + \mathrm{CAP}_g\,) \right) + \mathrm{MC}_{ft} + \widehat{p}_e\, E_f \right] + \widehat{\mu}_{ft} \ \ge \ 0, \quad \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \ \le \ \widehat{\mu}_{ft} \quad &\perp \quad \widehat{\mathrm{cap}}_f - \widehat{\bar{s}}_{ft} - \mathrm{CAP}_f \ \ge \ 0, \quad \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \ \le \ \widehat{p}_e \quad &\perp \quad \overline{E} - e_{\mathrm{NP}}(\widehat{p}_e) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t\, E_g\, \left(\widehat{\bar{s}}_{gt} + \mathrm{CAP}_g \right) \ \ge \ 0 \\
0 \ \le \ \widehat{p}_c \quad &\perp \quad \sum_{g \in \mathcal{F}} \widehat{\mathrm{cap}}_g - \underline{\mathrm{CAP}} \ \ge \ 0.
\end{aligned}
$$

By the same proof as that of Proposition 3, we can show that $\widehat{\bar{\mathbf{s}}} \neq 0$. Since

$$
\frac{\sum_{t \in \mathcal{T}} H_t\, (\, \bar{s}_{ft}^{\varepsilon_k} + \mathrm{CAP}_f\,)}{(\,\mathrm{cap}_f^{\varepsilon_k} + \varepsilon_k\,)} \ \le \ \frac{\sum_{t \in \mathcal{T}} H_t\, \mathrm{cap}_f^{\varepsilon_k}}{(\,\mathrm{cap}_f^{\varepsilon_k} + \varepsilon_k\,)} \ < \ \sum_{t \in \mathcal{T}} H_t
$$

it follows that the sequence $\left\{ \dfrac{\sum_{t \in \mathcal{T}} H_t\, (\, \bar{s}_{ft}^{\varepsilon_k} + \mathrm{CAP}_f\,)}{(\,\mathrm{cap}_f^{\varepsilon_k} + \varepsilon_k\,)} \right\}$ must have at least one accumulation point. With-

out loss of generality, we may assume that this sequence converges to a limit, say $\gamma_f \geq 0$. Note that

$$\gamma_f \;=\; \frac{\displaystyle\sum_{t\in\mathcal{T}} H_t\,(\,\widehat{\bar{s}}_{ft} + \mathrm{CAP}_f\,)}{\widehat{\mathrm{cap}}_f}, \quad \text{if the denominator is positive.}$$

Define, for each $f \in \mathcal{F}$,

$$\alpha_f \;\equiv\; \frac{(\overline{E} - E_{\mathrm{GF}})\,\widehat{R}_f\,\gamma_f}{\displaystyle\sum_{g\in\mathcal{F}} \widehat{R}_g \sum_{t\in\mathcal{T}} H_t\,(\,\widehat{\bar{s}}_{gt} + \mathrm{CAP}_g\,)}\,\widehat{p}_e.$$

It is easy to show that the following complementarity holds:

$$0 \;\leq\; \widehat{\mathrm{cap}}_f \;\perp\; \widehat{p}_c + F_f - \alpha_f\,\widehat{p}_e + \sum_{t\in\mathcal{T}} \widehat{\mu}_{ft} \;\geq\; 0, \quad \forall\, f \,\in\, \mathcal{F}.$$

With

$$\widehat{\alpha} \;\equiv\; \frac{(\overline{E} - E_{\mathrm{GF}})\,\widehat{p}_e}{\displaystyle\sum_{g\in\mathcal{F}} \widehat{R}_g \sum_{t\in\mathcal{T}} H_t\,(\,\widehat{\bar{s}}_{gt} + \mathrm{CAP}_g\,)},$$

it is easy to see that all conditions in (14) are satisfied.

5 An Application

To illustrate the results that can be obtained from our proposed models (as mentioned before, the NCPs (10) and (14) are the workhorses in the experiments), we consider a competitive power market at a single node with the following characteristics:

- Time periods: $T = 20$ periods per year, each $H_t = 438$ hours in length
- Demands: $d_t(p_t) = a_t - b_t p_t$, with $a_t = 500t$ and $b_t = t/2$
- Nonpower emission: $e_{\mathrm{NP}}(p_e) = 0$
- Generator types: $i = 1$ (coal steam), 2 (natural gas-fired combined cycle), and 3 (natural gas-fired combustion turbine)
- Minimal generation: $\mathrm{CAP}_1 = 0$ MW, $\mathrm{CAP}_2 = 0$ MW, and $\mathrm{CAP}_3 = 0$ MW
- Marginal costs: $\mathrm{MC}_1 = 20$ \$/MWh, $\mathrm{MC}_2 = 40$ \$/MWh, and $\mathrm{MC}_3 = 80$ \$/MWh
- Investment costs: $F_1 = 120,000$ \$/MW/yr, $F_2 = 75,000$ \$/MW/yr, and $F_3 = 50,000$ \$/MW/yr
- Firms' emission rates: $E_1 = 1$ ton/MWh, $E_2 = 0.35$ ton/MWh, and $E_3 = 0.6$ ton/MWh
- Total capacity requirement: $\underline{\mathrm{CAP}} = 11{,}000$ MW, if $\underline{\mathrm{CAP}} > 0$.

Thus, the demand function in each period is defined so that the peak load occurs during period 20, and load is proportional to t, if the same price is faced in each period. The demand curve parameters imply that the price intercept of the inverse demand curve is \$1000/MWh in each period; since equilibrium prices are usually under \$100/MWh, this means that demand is relatively inelastic at the equilibrium price. A capacity market is assumed that requires 10% more capacity than the peak demand of 10,000 MW that occurs if the price in the peak period was 0\$/MWh. Consumers are assumed to pay for capacity through a non-distorting customer charge; other assumptions are possible, such as allocation of capacity charges to peak energy prices, but are not explored here.

We introduce the following system performance measures:

- *Generation cost* (M\$/yr), total generation investment and fuel costs: $\displaystyle\sum_{f\in\mathcal{F}} \left(F_f \mathrm{cap}_f + \sum_{t\in\mathcal{T}} H_t MC_f s_f \right)$;

• *Social cost* (M$/yr), the cost of generation plus the cost of price-induced changes in energy consumption: PS + CS + GS, where PS is the equilibrium producer surplus, equal to the sum over all firms of the objectives in (1); CS is the equilibrium consumer surplus: $\sum_{t\in\mathcal{T}} H_t \left(\int_0^{d_t(p_t)} \pi_t(x)dx - d_t(p_t)p_t \right)$; and GS is the auction or grandfathering surplus, equal to the economic rent accruing to the original owners of grandfathered allowances (or, equivalently, the revenue received by the government if it instead auctioned those allowances). However, we exclude environmental costs from this performance measure;

• *Consumer payments* (M$/yr): $p_c \sum_{f\in\mathcal{F}} \text{cap}_f + \sum_{t\in\mathcal{T}} d_t(p_t)p_t$; and

• *Capacity factor*, the ratio of annual generation to potential generation: $\dfrac{\sum_{t\in\mathcal{T}} H_t s_{ft}}{\sum_{t\in\mathcal{T}} H_t \text{cap}_f}$.

In the computations, the NCPs (10) and (14) and the linear complementarity problem (23) below are solved by the PATH solver available on the NEOS server (http://neos.mcs.anl.gov/neos/solvers/index.html).

5.1 The base run

To provide a basis for comparison of the two emission rules, we derive an equilibrium solution in the absence of a CO_2 limit; i.e., with $\overline{E} = \infty$; thus the constraint (3) is absent and $p_e = 0$. Such an equilibrium, which we call the *(emission) unconstrained solution*, is the solution of the following linear complementarity problem:

$$
\begin{array}{lcll}
0 \le \bar{s}_{ft} & \perp & H_t \left[-\pi_t \left(\sum_{g\in\mathcal{F}} (\bar{s}_{gt} + \text{CAP}_g) \right) + \text{MC}_{ft} \right] + \mu_{ft} \ge 0, & \forall\, (f,t) \in \mathcal{F}\times\mathcal{T} \\
0 \le \mu_{ft} & \perp & \text{cap}_f - \bar{s}_{ft} - \text{CAP}_f \ge 0, & \forall\, (f,t) \in \mathcal{F}\times\mathcal{T} \\
0 \le \text{cap}_f & \perp & -p_c + F_f - \sum_{t\in\mathcal{T}} \mu_{ft} \ge 0, & \forall\, f \in \mathcal{F} \\
0 \le p_c & \perp & \sum_{g\in\mathcal{F}} \text{cap}_g - \underline{\text{CAP}} \ge 0. &
\end{array}
\tag{23}
$$

When the total capacity requirement $\underline{\text{CAP}}$ is 11,000 MW, capacity cap_f of coal, combined cycle, and combustion turbines equal 7329 MW, 1628 MW, and 2042 MW, respectively. The bulk (94%) of the energy is obtained from coal plants, with combined cycle facilities providing nearly all of the remainder. Combustion turbines are built primarily to meet the capacity market requirement. Emissions amount to 47.4 Mtons/yr. The total cost of generation is 2049 M$/yr, which also equals consumer payments for energy, under the zero profit free-entry assumption. Energy prices p_t equal the marginal cost of generation in each period, varying from $20 to $80/MWh, depending on the marginal source of energy; meanwhile, the capacity market price p_c equals the annual cost of combustion turbine capacity, 50,000 $/MW/yr. The quantity-weighted power price is 46.1 $/MWh, of which capacity payments make up 27%. The sum of consumer surplus and producer surplus is 20911 M$/yr.

5.2 Comparative results for the allowance allocation rules

Table 1 reports the main system performance measures for a series of experiments representing alternative assumptions regarding: the type of the contingent emission allocation, i.e., rules (I) and (II) as well as the base run where $\overline{E} = \infty$; the presence or absence of a capacity market; and the presence or absence of minimum output levels (in the form of a min-run capacity constraint) for coal plants only. For each set of

assumptions, results are presented for CO_2 limits of 20 (a severe restriction) and 40 (a mild restriction) Mtons/yr, and for three cases of 0%, 50% or 100% grandfathering of allowances (corresponding to 100%, 50%, and 0% of allowances granted to new generating plants; i.e., $E_{GF}/\overline{E} = 0, .5, 1$, respectively). We take R_f in (7) and $\widehat{R}_f$ in (6) to be both equal to E_f/E_1.

With the presence of a capacity market, runs 1–6 show the results for the potential emission rule and runs 7–12 show the results for the actual emission rule. For the purpose of the comparison, we focus on the increase in generation cost, social cost (equal to the loss of social surplus), and consumer payments relative to the (emission) unconstrained solution obtained in Subsection 5.1. We do not report producer surplus, because it is by assumption zero in the free entry solutions. For ease of comparison, the increases are expressed as a percentage of the production cost of the unconstrained solution (2049 \$M/yr), which also equals the consumer payment in that solution. The three percentage increases are calculated, respectively, as:

- relative generation cost increase: 100% (generation cost - 2049 M\$/yr)/2049 M\$/yr;
- relative social cost increase: 100% [20911 M\$/yr - (PS + CS + GS)]/2049 M\$/yr; and
- relative consumer payments increase: 100% (consumer payments - 2049 M\$/yr)/2049 M\$/yr.

The imposition of the CO_2 constraint causes social surplus to decrease relative to the unconstrained value of 20911 M\$/yr. This decrease is not identical to the change in generation cost because of changes in energy consumption that are caused by shifts in energy prices. If demand elasticity was instead zero, then the change in social cost would be the same as the change generation cost. The 100% grandfathered cases are the same for both contingent allocation rules because, of course, that level of grandfathering means no allowances are allocated to new investment. In the case where all allowances are instead allocated to new plants by the potential emission rule (Runs 1,4), we see that, like the actual emission rule, the investment is greatly distorted and costs are much higher than if allowances are completely grandfathered (or auctioned). In fact, the distortion is worse, because it is possible for new generators to receive free allowances even if they do not generate power. The allowances given to new investment are sufficiently valuable so that it is worthwhile to build combustion turbines, even though they don't operate. In the 20 Mton/yr limit case, 95% of the combustion turbine capacity is never used, and is built just to collect free allowances. Unexpectedly, the distortion is much worse in the mild (40 Mton/yr) limit case, making the total cost of compliance almost as high as for the 20 Mton/yr case. This confirms that it should not be assumed that the risk of distortion is less if CO_2 limits are less severe.

The next set of alternative assumptions addresses the effect of assuming an energy-only market in which there is no separate market for capacity. In this case (Runs 13–24), energy prices rise much higher during peak periods to ensure that consumer demand does not exceed available capacity. There are extensive debates regarding the pros and cons of capacity markets (e.g., see [10]), which we do not consider here. Instead we merely consider the interaction of capacity and emissions markets. The main effect observable in comparing Runs 13-24 with Runs 1-12 is in the case of the actual emission rule. Cost increases due to investment distortion are somewhat greater without the capacity market, especially in the 40 Mton/yr limit, and there is less investment. The decreased investment is most dramatic for combustion turbines, which are not built at all under the actual emission rule. In contrast, under the potential emission rule, the costs and generation mixes under 0% grandfathering (all allowances given to new investment) are completely unaffected by the presence of a capacity market. This is because the potential emission form of the contingent rule results in overinvestment such that the capacity constraint is not binding.

The final set of alternative assumptions we consider is the imposition of a min-run constraint on coal plants, since in reality they cannot really be cycled in the way they are in the solutions that have low coal capacity factors. We assume that the output of coal plants cannot be reduced below 35% of their capacity; as a result, in periods where that constraint is binding, energy prices can actually be negative.

This phenomenon is occasionally observed during low demand periods in real power markets. The min-run assumption significantly raises costs in all scenarios, particularly so in the no-CO_2 constraint case, because it has the most coal capacity. As a result, the cost impact of imposing a CO_2 limit is reduced by well over half. This can be seen by comparing Runs 25-27 with Runs 1-3, which, aside from omitting the min-run constraint, make the same assumptions. The cost and generation mix distortions resulting from using a contingent allocation rule (in this case, the potential emission rule) is diminished somewhat, but remains large. The cost of complying with the 20 Mton/yr limit is more than doubled compared to the equilibrium under 100% grandfathering or auctioning of emissions.

5.3 Detailed results for the actual emissions rule

To take a more detailed look at the effect of grandfathering, we solved a set of problems with the percentage of allowances grandfathered varying from 0% to 100% in increments of 10%, assuming the presence of a capacity markets. The results are plotted in Figures 2–5. We show the results for the actual emission rule under the two levels of CO_2 restrictions: 20 Mtons/yr (Figures 2 and 4) and 40 Mtons/yr (Figures 3 and 5). In the horizontal axis of each figure, we vary the fraction of emissions that are grandfathered from 0% to 100%. Thus, the equilibrium on the far left allocates all allowances to new investment by the actual emission rule, which means that freely granted allowances exactly equals actual emissions for each plant type, so that generators pay nothing, on net, for their emissions. On the other hand, at the right-hand extreme, all allowances are grandfathered or auctioned, so that new capacity has to pay for 100% of their emissions.

Two sets of results are shown for each emissions cap: Figures 2 and 3 show the effect of the different policies upon three categories of costs and the price of allowances p_e, while Figures 4 and 5 show how the mix of generation investment and the operation of coal and combined cycle plants are affected. (The results for the three cases of 0%, 50% and 100% grandfathering of allowances are already reported in Runs 7–12 in Table 1.)

Note that consumer payments are in excess of generation investment and fuel costs if some of the allowances are grandfathered (Figures 2 and 3). This is because in the zero profit equilibrium, the revenues that generators receive have to cover not only investment and fuel costs, but also the purchase of grandfathered (or auctioned) allowances. From a social cost point of view, however, the expense associated with such allowances is just an income transfer from generators (and thus consumers) to the owners of the grandfathered allowances (or the government, if instead those allowances are auctioned). Figure 3 shows that under complete grandfathering, the 40 Mton/yr limit has a very small social cost: 19 M$/yr, or less than 1% of total investment and operating cost. However, the stricter 20 Mton/yr limit is much more expensive, 395 M$/yr, which is almost 20% of the unconstrained generation cost, see Figure 2. Comparing the right-hand bars of Figures 4 and 5, we see that the cost difference arises because much more gas-fired generation is required in order to meet the stricter standard.

Under both limits, costs increase further if instead some or all allowances are freely given to new entry under the actual emission rule. The distortions, as measured by generation or social cost, are mild until the fraction of grandfathered allowances falls below 80%, and then increase rapidly as that fraction falls further. In the extreme case of all allowances being freely allocated, the costs of the CO_2 constraint are greatly inflated. For instance, rather than 19 M$/yr and 395 M$/yr under the loose and tight CO_2 constraints, respectively, the social costs of the constraint rise to 242 M$/yr and 512 M$/yr, respectively (12% and 25%, respectively, of the unconstrained generation cost). Thus, the free allocation of allowances to new entry has inflated the cost of meeting the CO_2 constraint by more than an order of magnitude under the loose CO_2 constraint.

Although intuition might suggest that the inefficiency would be less under the looser constraint, it is actually about twice as large (242 minus 19, as opposed to 512 minus 395) as under the tight constraint. The reasons for these distortions are revealed by Figure 4 and Figure 5. As the fraction of allowances

that are grandfathered decreases, the price of emissions allowances increases. Surprisingly, this results in greater investment in coal-fired capacity but paradoxically relatively little change in generation from such facilities. What is happening is that allowance prices climb to the point ($31/ton) where natural gas plants are cheaper to run, on the margin, than coal plants, when the opportunity cost of allowances is factored into the cost. The dispatch order is then reversed compared to the 100% grandfathered case, with combined cycle plants being base loaded and coal plants being cycled on and off. This is reflected in the shifts in capacity factors shown in the figures. Base loading the combined cycle plants greatly increases their capacity factors, while coal plant output falls to as little as 30% of their maximum possible production. The allowances are so valuable that the net capacity cost of coal plants, including the value of the free allowances they are given, falls below the capital cost of peaking turbines, which no longer enter the market. This is reflected in the price of capacity dropping below the turbine capital cost.

The large increases in social cost resulting from contingent allocation of allowances primarily reflect these distortions in investment. For instance, in the 40 Mton/yr limit case, the increase in capital costs arising from the larger investments in coal facilities at the expense of cheaper gas-fired plants amounts to 234 M$/yr, out of the total increase in social cost of 242 M$/yr relative to the least-cost way of achieving that standard. The results indicate that energy prices are lower under 0% grandfathering than under 100% grandfathering. This is partially consistent with the conjecture of the Netherlands Bureau for Economic Policy Analysis [14] that much of the value of freely-granted allowances is passed back to consumers in the long run. However, not all that value is returned; much of it is instead eaten up by the investment distortions. If the economic rent associated with grandfathered allowances could be returned to consumers either via tax reductions or other mechanisms, then consumers would generally be much better off with 100% grandfathering than with contingent allocation of allowances.

6 Conclusion

In this paper, we have presented complementarity problem formulations for the analysis of alternative emissions allowance allocation systems in electric power markets. Existence of solutions was proven under mild conditions. An example illustrates the potential for investment distortion arising from allocation rules that give allowances to new capacity. Future work could address formulation of more realistic models including, for instance, transmission or carbon sequestration alternatives; parameterization based on actual markets; extension to other allowance allocation systems, such as output-based allocation [4]; and representation of interlinked markets in which different markets are subject to different rules, as is presently the case in the European Union.

References

[1] M. Ahman and K. Holmgren (2006). New entrant allocation in the Nordic energy sectors: Incentives and options in the EU ETS. *Climate Policy*, in press.

[2] M. Bartels, and F. Musgens (2006). Do technology specific CO_2-allocations distort investments? Institute of Energy Economics, University of Cologne, Germany.

[3] D. Burtraw, K. Palmer, R. Bharvirkar, and A. Paul (2001). The effect of allowance allocation on the cost of carbon emission trading. RFF DP 01-30, Resources for the Future, Washington, DC.

[4] D. Demailly and P. Quirion (2006). CO_2 abatement, competitiveness and leakage in the European cement industry under the EU ETS: grandfathering versus output-based allocation. *Climate Policy* **6** 93–113.

[5] Y. Dissou (2005). Cost-effectiveness of the performance standard system to reduce CO_2 emissions in Canada. *Resource and Energy Economics* **27** 187-207.

[6] F. Facchinei and J.S. Pang (2003). *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Volumes I and II, Springer-Verlag, New York.

[7] C. Fischer (2001). Rebating environmental policy revenues: Output-based allocations and tradable performance standards , RFF DP 01-22, Resources for the Future, Washington, DC.

[8] C. Fischer and A. Fox (2004). Output-based allocation of emissions permits for mitigating tax and trade interactions. RFF DP 04-37, Resources for the Future, Washington, DC.

[9] M. Grubb and K. Neuhoff (2006). Allocation and competitiveness in the EU emissions trading scheme: Policy overview. *Climate Policy* **6**(1) 7–30.

[10] B.F. Hobbs, M.C. Hu, J. Inon, M. Bhavaraju, and A. Paul (2007). A dynamic analysis of a demand curve-based capacity market proposal: the PJM reliability pricing model. *IEEE Transactions on Power System* **22**(1) 3–11.

[11] P. Linares, F.J. Santos, M. Ventosa, and L. Lapiedra (2006). Impacts of the European emissions trading scheme directive and permit assignment methods on the Spanish electricity sector. *The Energy Journal* **27**(1) 79–98.

[12] F. Matthes, V. Graichen, and J. Repenning (2005). The environmental effectiveness and economic efficiency of the European Union Emissions Trading Scheme: Structural aspects of allocation. [Prepared for the World Wildlife Federation, Institute for Applied Ecology, Berlin.] Netherlands Bureau for Economic Policy Analysis, Ministry of Economic Affairs. Explanation of CPB Vision on Relationship Emissions Trading - Power Prices. Unpublished memo, The Hague.

[13] C.B. Metzler, B.F. Hobbs, and J.S. Pang (2003). Nash-Cournot equilibria in power markets on a linearized DC network with arbitrage: formulations and properties. *Networks and Spatial Economics* **3**, 123–150.

[14] Netherlands Bureau for Economic Policy Analysis (2005). Explanation of CPB vision on relationship emissions trading–power prices. Unpublished memo, The Hague.

[15] K. Neuhoff, M. Ahman, R. Betz, J. Cludius, F. Ferrario, K. Holmgren, G. Pal, M. Grubb, F. Matthes, K. Rogge, M. Sato, J. Schleich, J. Sijm, A. Tuerk, C. Kettner, and N. Walker (2006). Comparison of national allocation plans for the the period 2008-2012. EPRG Working Paper, Electric Power Research Group, Cambridge University.

[16] K. Neuhoff, M. Grubb, and K. Keats (2005). Impact of allowance allocation on prices and efficiency. CWPE 0552 and EPRG 08, Electric Power Research Group, Cambridge University.

[17] K. Neuhoff, K.K. Martinez, and M. Sato (2006). Allocation, incentives and distortions: The impact of EU ETS emission allowance allocations to the electricity sector. *Climate Policy* **6**(5) 73–91.

[18] K. Palmer, D. Burtraw, and D. Kahn (2006). Simple rules for targeting CO_2 allowance allocations to compensate firms. RFF DP 06-28, Resources for the Future, Washington, DC.

[19] J.S. Pang, and B.F. Hobbs (2004). Spatial oligopolistic equilibria with arbitrage, shared resources, and price function conjectures. *Mathematical Programming, Series B* **101**, 57–94

[20] J. Sijm (2006). EU ETS allocation: Evaluation of present system and options beyond 2012. *Zeitshrift für Energiewirtschaft* **30**(4) 285–292.

[21] J. Sijm, K. Neuhoff, and Y. Chen (2006). CO_2 cost pass-through and windfall profits in the power sector. *Climate Policy* **6**(5) 49–72.

[22] Y. Smeers and A. Ehrenmann (2006). Free allowances and investments in a CO_2 constrained restructured market. INFORMS National Meeting, Pittsburgh, PA, November.

[23] T. Sterner, and A. Muller (2006). Output and abatement effects of allocation readjustment in permit trade. RFF DP 06-49, Resources for the Future, Washington, DC.

[24] T. Tietenberg (2006). *Emissions Trading-Principles and Practice*. Resources for the Future Press, Washington, DC.

Table 1. Summary of Model Results

Capacity Market (MW)	Min Output Level	Contingent Allocation System	Run	CO2 Limit(Mton/yr)	% Grandfathered	% Increase Relative to Base Run			Capacity(MW)			% Capacity Factor			p_e ($/ton)	p_c ($/ton)
						Generation Cost	Social Cost	Consumer Payments	Coal	Comb. Cycle	Comb. Turbine	Coal	Comb. Cycle	Comb. Turbine		
11000	None	Potential	1	20	0%	28.7%	31.6%	28.7%	3088	5540	5361	27.6%	73.2%	0.3%	34.35	0
			2		50%	16.9%	19.4%	31.3%	863	7747	2390	97.8%	52.6%	0.9%	29.54	14617
		Emission	3		100%	17.6%	19.3%	39.5%	852	8084	2064	97.8%	51.1%	0.3%	22.45	50000
			4	40	0%	26.5%	29.2%	26.5%	8094	559	12155	53.9%	99.2%	0.1%	32.46	0
			5		50%	-0.2%	2.5%	29.8%	6803	1729	2468	64.0%	32.0%	1.0%	30.77	8462
		Rule	6		100%	0.1%	0.9%	21.6%	6076	2871	2053	71.1%	23.9%	0.5%	11.01	50000
		Actual	7	20	0%	23.8%	25.0%	23.8%	2459	8541	0	33.7%	48.6%	N/A	30.77	29115
			8		50%	20.1%	21.2%	35.1%	1600	7889	1511	51.8%	52.7%	0.0%	30.77	50000
		Emission	9		100%	Same results as Run 3										
			10	40	0%	10.6%	11.8%	10.6%	10140	860	0	42.8%	74.6%	N/A	30.77	4593
			11		50%	5.0%	6.2%	35.0%	8359	1211	1430	51.9%	53.0%	0.0%	30.77	50000
		Rule	12		100%	Same results as Run 6										
		Base Run		No Limit		2049	20911	2049	7329	1628	2042	65.10%	17.65%	0.59%	0.00	50000
None	None	Potential	13	20	0%	Same as results as Run 1										N/A
			14		50%	22.8%	23.8%	39.1%	1976	6557	290	43.2%	61.9%	5.6%	30.77	
		Emission	15		100%	18.7%	20.6%	42.4%	886	7601	0	97.7%	53.3%	N/A	22.45	
			16	40	0%	Same as results as Run 4										
			17		50%	5.1%	5.8%	38.2%	8038	553	301	54.3%	99.3%	5.0%	31.43	
		Rule	18		100%	-0.1%	0.9%	23.2%	6160	2365	0	70.7%	25.7%	N/A	11.01	
		Actual	19	20	0%	25.3%	24.8%	25.3%	2226	6851	0	37.8%	60.2%	N/A	30.77	N/A
			20		50%	21.6%	22.5%	37.9%	1606	7109	0	53.4%	57.3%	N/A	30.77	
		Emission	21		100%	Same results as Run 15										
			22	40	0%	17.1%	15.9%	17.1%	9752	807	0	44.5%	79.5%	N/A	30.77	
			23		50%	4.2%	4.9%	36.7%	7859	889	0	55.6%	63.4%	47.3%	30.77	
		Rule	24		100%	Same results as Run 18										
		Base Run		No Limit		1893	21020	1893	7329	1232	0	65.10%	19.79%	N/A	0.00	N/A
11000	Coal Only	Potential	25	20	0%	13.4%	13.5%	13.4%	1441	7018	5811	58.5%	57.8%	0.6%	30.77	0
		Emission	26		50%	5.1%	5.0%	18.1%	863	7747	2390	97.8%	52.6%	0.9%	29.54	14617
		Rule	27		100%	5.8%	4.9%	25.5%	852	8084	2064	97.8%	51.1%	0.3%	22.45	50000
		Base Run		No Limit		2278	20628	2278	2287	6671	2042	0.9%	0.4%	0.0%	0.00	50000

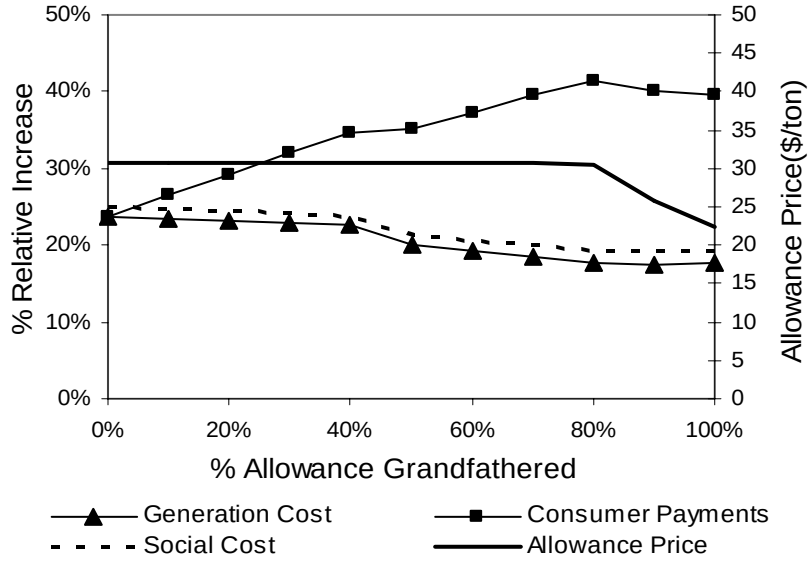

Figure 2: Cost and price comparison (20 Mton limit)

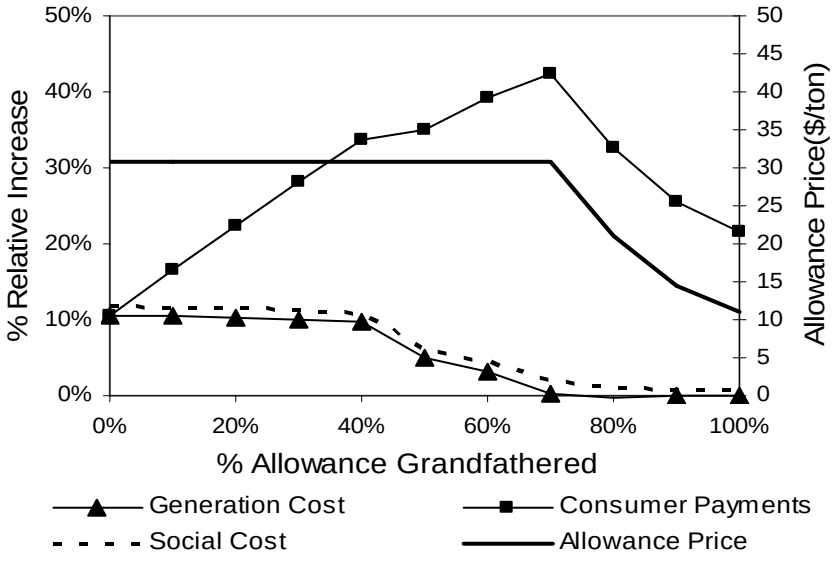

Figure 3: Cost and price comparison (40 Mton limit)

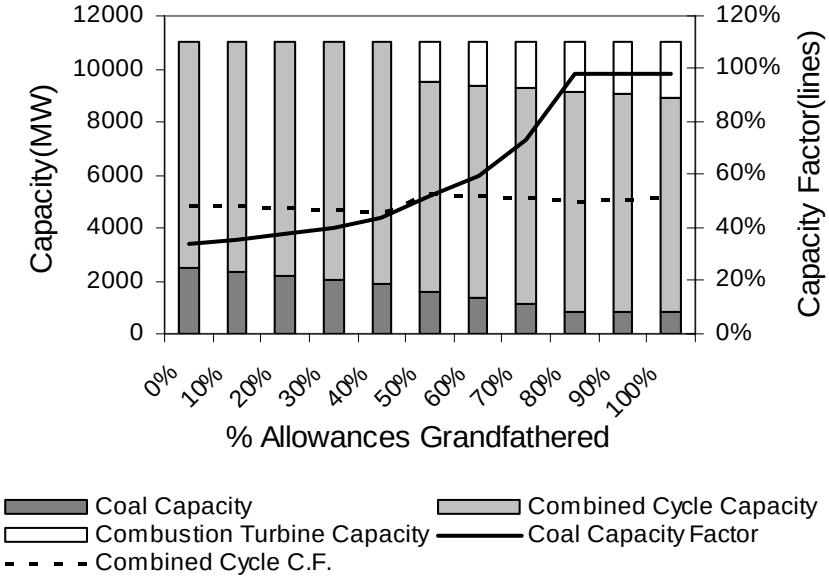

Figure 4: Capacity and capacity factor comparison (20 Mton limit)

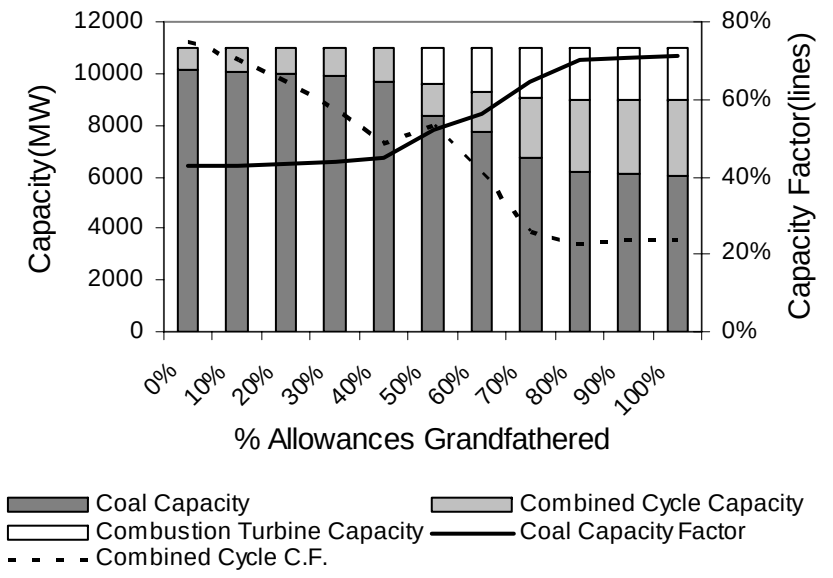

Figure 5: Capacity and capacity factor comparison (40 Mton limit)