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

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

August 24, 2009

Subject classification: Equilibrium programming: model properties and applications. Economics: electricity and emissions markets.

Area of review: Environment, energy, and natural resources.

History: Received May 2007; revision received August 2008, May 2009; accepted August 2009.

Abstract

Carbon dioxide allowance trading systems for electricity generators are in place in the European Union and in several U.S. states. An important question in the design of such systems is how allowances are to be initially allocated: by auction, by giving away fixed amounts (grandfathering), or by allocating based on present or recent output, investment, or other decisions. The latter system can bias investment, operations, and product pricing decisions, and increase costs relative to the other systems. A nonlinear complementarity model is proposed for investigating the long-run equilibria that would result under alternative systems for power markets characterized by time varying demand and multiple technology types. Existence of equilibria is shown under mild conditions. Solutions for simple systems show that allocating allowances to new capacity based on fuel use or generator type can yield large distortions in capacity investment, can invert the operating order of power plants, and inflate consumer costs. The distortions can be smaller for tighter CO_2 restrictions, and are somewhat mitigated if there is also a market for electricity capacity or minimum-run restrictions on coal plants. Distortions are also less if allowances are allocated to plants in proportion to sales rather than capacity.

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 [25].

*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.

†ISO-New England, One Sullivan Rd., Holyoke, MA 01040, U.S.A. Email: jzhao@iso-ne.com. Earlier drafts of the paper were written when the first and third authors were 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.

‡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@illinois.edu.

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. Although there is no US federal CO_2 reduction requirement as of 2008, several states have initiated their own cap-and-trade systems for CO_2, notably 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 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 were blamed upon the introduction of the ETS [22].

There are many aspects of the design of a cap-and-trade system that can affect its economic efficiency and costs to 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. In principle, emission allowances can either be (1) *auctioned*, with polluters purchasing them from their regulators; (2) *grandfathered* to polluters in fixed amounts free of charge based perhaps on unalterable historical decisions by polluters; or (3) allocated based on the firms' present or recent actions, i.e., *contingent allocation rules*. 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 [13]. 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 [21]. "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 social costs of pollution, as our focus is on economic efficiency effects given a fixed amount of pollution (i.e., the cap) rather than on tradeoffs between costs and pollution levels.

As an example of how allocation rules can distort production decisions, if a contingent allocation system (also called "output-based" or "input-based" allocation) gives future allowances in a way that depends on present generator decisions, incentives to deviate from least-cost investment mixes and operation are introduced. For instance, in the EU ETS, during Phase I (2005-2007), it was announced that allowance allocations in Phase II (post- 2007) would depend in part upon Phase I emissions, arguably providing an incentive to expand pollution in the earlier 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 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. Furthermore, if dirtier new sources are allocated more allowances per unit capacity or unit output, as was done in at least eight EU countries [16], then technology choices can be skewed. Finally, if different jurisdictions in the same power market have different allocation rules, as in the EU ETS, then 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 [18, 24]. This support is in spite of numerous modeling studies that have compared the relative efficiency of such schemes compared to auctions and grandfathering, and 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 auctioning or grandfathering, 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 improve welfare relative to auctions and grandfathering [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 or in proportion to production, and systems in which the initial allocation of allowances does not depend on present or future capital or operating decisions (either auction or grandfathering). 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)" [15]. We do not address other important issues concerning the design of emissions allocation systems. Some of these include [21]: 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 or floors, designed to stabilize prices; the value of "banking" schemes to buffer interannual variations in emissions; and 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 [17] 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 [19] 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, 17, 24] to gain more allowances later. But Neuhoff, Martinez, and Sato [18] point out that such 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 [24]. The simulations of this paper show a similar result. Two papers [17, 18] 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 [23] looks at the complications introduced by a 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. Such rules can include caps that limit the price of energy during periods of shortage, or the absence of scarcity pricing mechanisms that allow prices to rise above marginal cost in order to motivate demand reductions. 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 [23] which in addition considered capacity market failures. The most comparable model is that of Section 4 of [17]. Like that model, we consider a long-run, free-entry equilibrium among more than one technology in which all capacity is variable and there are no incumbent generators. We also share their implicit assumptions that long-run contract markets and short-run spot markets are arbitraged; that generators are price takers (although firms exhibiting market power can be handled easily in our framework); and 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 [17] 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 that can also 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 [21].

However, a price is paid for these added complexities; 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, 17, 24]. 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 the *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 three different contingent allocation rules. The first two rules allocate allowances to new capacity irrespective of how individual plants are operated. The first (Rule I) gives allowances in proportion to potential emissions and the other (Rule II) allocates them in proportion to average actual total emissions in the market for each capacity technology. The third (Rule III) instead allocates allowances in proportion to plant production rather than capacity. Each of those rules can be parameterized to change the relative amounts of allowances for different technologies. 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, with the allocation rules based on capacity (I, II) rather than sales (III) having a much greater effect. 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

Figure 1 depicts the various components of the market and their interrelations. The model is a static model of generator and customer decisions under perfect competition. The equilibrium is meant to represent a static long-run steady-state in which all capacity is variable, so all plants in the market are eligible for allowances that are assigned to new entry. If any allowances have been grandfathered to existing capacity at the time of establishment of the emissions trading program, we assume that this capacity has been entirely retired. However, we also assume that the owners of that retired capacity continue to receive any grandfathered allowances and then sell them to owners of operating capacity. For simplicity of presentation, each firm is assumed to own only one type of generating capacity; more general models are easily formulated but more complex to present. Generation firms participate in up to four types of markets: energy, fuels (not explicitly modeled here, but see [20]), capacity (considered in some scenarios), and emissions allowances. Because demand varies over the day and year and because electric power is difficult to store, energy markets at different times during the year are separate markets with distinct prices. Rather than consider 8760 different energy markets (one per hour), we aggregate hours with similar demands into single periods t, with T distinct periods per year. The variation in demand and prices means that, in equilibrium, there will be a mix of both capital-intensive plants that operate most hours and fuel-intensive but cheap to build plants that generate only during peak demand periods. Shifts in the mix of investment among these technology types is one source of inefficiency of different allowance allocation systems, and so multiple periods need to be represented explicitly in this model. A capacity market is represented because it is present in many power markets. Capacity markets are designed to ensure achievement of a prespecified level of capacity reserves for reliability reasons. Their presence directly affects the amount and, sometimes, the types of capacity that are built, and this effect is altered by the choice of allowance allocation system, as seen in our simulations. Finally, the emissions market either provides producers with allowances free of charge according to the assumed rule, or charges for allowance purchases. Generators with more allowances than they need can sell that excess back to the market.

We summarize the notation used in the model formulation; first the parameters, 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.

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

$\mathcal{F}$	Set of firms
$\mathcal{T}$	Set of time periods $\equiv \{1, \cdots, T\}$
MC_{ft}	Marginal cost for firm f in period t, excluding cost of emission allowances (\$/MWh)
E_f	Emission rate for firm f (tons/MWh)
F_f	Annualized investment cost of firm f's capacity (\$/MW/yr)
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. Nine EU countries differentiate among generation technologies by, in effect, assigning different values of this ratio [16].
$\overline{E}$	Total emission allowances supply (tons/yr): $\overline{E} > E_A$
E_A	Amount of emission allowances that are auctioned (or grandfathered)(tons/yr)
H_t	Hours in period t (hr/yr), here assumed to be $8760/T$
χ	Unit converter = 1 MW^2 yr/\$
$\underline{CAP}$	Total capacity requirement (MW)

$\overline{E} > E_A$ 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 $d_t(\bullet)$; (\$/MWh)
$e_{NP}(\bullet)$	Allowance sales to the nonpower sector from the power sector as a function of the allowance price, which is nonincreasing (tons/yr)

Variables:

p_t	$= \pi_t\left(\sum_{g\in\mathcal{F}} s_{gt}\right)$: Energy price during period t (\$/MWh)
p_e	Emission allowance price (\$/ton)
p_c	Capacity price (\$/MW/yr)
α_f	Emission allowance allocation parameter for firm f (tons/MW for Rules I,II or tons/MWh for Rule III)
s_{ft}	Energy sold by firm f in period t (MW)
cap_f	Capacity for firm f (MW)
μ_{ft}	Dual variable associated with firm f's capacity constraint in period t (\$/MW/yr)

The basic model has three main components: (a) firms' profit maximization problems, (b) 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 Generating firms' optimization problems

First we consider the optimization problem under Rules I and II (where allowances are allocated to capacity) and then present the problem for Rule III (allowances allocated directly to sales). 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

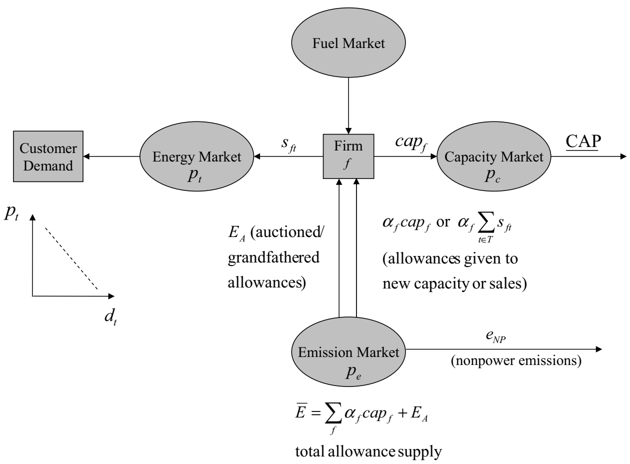

Figure 1: Market structure

function is revenue less cost, to determine its capacity cap_f and sales s_{ft}:

$$\begin{array}{ll} \underset{cap_f,\,(s_{ft})_{t\in\mathcal{T}}}{\text{maximize}} & \sum_{t\in\mathcal{T}} H_t\,(\,p_t - MC_{ft} - p_e\,E_f\,)\,s_{ft} + (\,p_c + p_e\alpha_f - F_f\,)\,cap_f \\ \text{subject to} & 0 \le s_{ft} \le cap_f, \quad \forall\, t \in \mathcal{T} \end{array} \tag{1}$$

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. This is consistent with the phrasing of some phase II emissions allocation rules in the ETS. Neuhoff et al. [16] report that fully half of the EU countries are planning to allocate at least some allowances in proportion to new capacity according to this formula, independent of the actual output of the individual plant.

The problem (1) is a linear program whose optimality conditions are:

$$\begin{array}{llll} 0 \le s_{ft} & \perp & H_t\,(\,-p_t + MC_{ft} + p_e\,E_f\,) + \mu_{ft} \ge 0, & \forall\, t \in \mathcal{T} \\ 0 \le \mu_{ft} & \perp & cap_f - s_{ft} \ge 0, & \forall\, t \in \mathcal{T} \\ 0 \le cap_f & \perp & -p_c - p_e\,\alpha_f + F_f - \sum_{t\in\mathcal{T}} \mu_{ft} \ge 0, & \end{array} \tag{2}$$

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

For the case of the sales-based Rule III, the above equations are modified thus so that the firms are awarded free allowances based on their sales and not capacity investment:

$$\begin{array}{ll} \underset{cap_f,\,(s_{ft})_{t\in\mathcal{T}}}{\text{maximize}} & \sum_{t\in\mathcal{T}} H_t\,(\,p_t - MC_{ft} - p_e\,E_f\, + p_e\alpha_f)\,s_{ft} + (\,p_c - F_f\,)\,cap_f \\ \text{subject to} & 0 \le s_{ft} \le cap_f, \quad \forall\, t \in \mathcal{T} \end{array} \tag{3}$$

The KKT conditions are:

$$\begin{array}{llll} 0 \le s_{ft} & \perp & H_t\,(\,-p_t + MC_{ft} + p_e\,E_f\, - p_e\alpha_f) + \mu_{ft} \ge 0, & \forall\, t \in \mathcal{T} \\ 0 \le \mu_{ft} & \perp & cap_f - s_{ft} \ge 0, & \forall\, t \in \mathcal{T} \\ 0 \le cap_f & \perp & -p_c + F_f - \sum_{t\in\mathcal{T}} \mu_{ft} \ge 0, & \end{array} \tag{4}$$

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

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

and an additional term corresponding to the derivative of the price function $\pi_t(\bullet)$ 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 [14, 20], as well as linear constraints, such as a "min-run capacity constraint" that is of the form $s_f \geq \gamma_f cap_f$ 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), (3) and their equivalent optimality conditions (2), (4) from now on.

2.2 Market clearing conditions

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

$$0 \,\leq\, p_e \,\perp\, \overline{E} - e_{NP}(p_e) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, s_{gt} \,\geq\, 0, \tag{5}$$

which stipulates that the 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 $e_{NP}(p_e)$ represents the effective demand for allowances from other sectors [12].

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

$$0 \,\leq\, p_c \,\perp\, \sum_{g\in\mathcal{F}} cap_g - \underline{CAP} \,\geq\, 0, \tag{6}$$

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 three contingent allocation rules for emissions allowances satisfy the condition that the amount of allowances available for allocation equals the amount allocated to capacity or to sales, as the case may be:

$$\begin{aligned} \overline{E} - E_A &= \sum_{f\in\mathcal{F}} \alpha_f\, cap_f && \text{Rules I, II} \\ \overline{E} - E_A &= \sum_{(f,t)\in\mathcal{F}\times\mathcal{T}} H_t\, \alpha_f\, s_{ft} && \text{Rule III} \end{aligned} \tag{7}$$

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

I The capacity-based potential emission (or input) rule, where

$$\alpha_f \, cap_f \;=\; \frac{R_f \, cap_f}{\displaystyle\sum_{g \in \mathcal{F}} R_g \, cap_g} \, (\, \overline{E} - E_A \,), \quad \forall \, f \, \in \, \mathcal{F}, \tag{8}$$

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 (7). 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 capacity-based actual emission (or output) rule which allocates allowances in proportion to a firm's capacity, but the proportion is based on the actual tonnage of emissions, where

$$\alpha_f \, cap_f = \frac{\widehat{R}_f \displaystyle\sum_{t \in \mathcal{T}} H_t \, s_{ft}}{\displaystyle\sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t \, \widehat{R}_g \, s_{gt}} \, (\, \overline{E} - E_A \,), \quad \forall \, f \, \in \, \mathcal{F}; \tag{9}$$

provided that the denominator is positive; or equivalently, $\alpha_f = \widehat{\alpha} \widehat{R}_f \frac{\sum_{t \in \mathcal{T}} H_t s_{ft}}{cap_f}$ (if $cap_f > 0$) for a common variable $\widehat{\alpha}$, due to the same emission balancing constraint (7). Although firms see this as a capacity-based rule (so they believe that doubling of capacity will double the allowances allocated), in equilibrium, the allocation will depend on the energy production by the firm's technology. If the $\widehat{R}_f$ are equal for all f, then this rule implicitly allocates allowances in proportion to sales, although unlike Rule III, the firm views the allocation as being proportional to capacity. However, if $\widehat{R}_f$ instead is the emissions rate per MWh, then allowances are implicitly allocated in proportion to emissions. In that case, in contrast to rule I, this rule ensures that if, say, a technology emits 75% of the CO_2, it receives 75% of the allowances $\overline{E} - E_A$ that are allocated to new capacity.

III The sales-based output rule which allocates allowances in proportion to a firm's energy production, where

$$\alpha_f \sum_{t \in \mathcal{T}} H_t \, s_{ft} \;=\; \frac{\widehat{R}_f \displaystyle\sum_{t \in \mathcal{T}} H_t \, s_{ft}}{\displaystyle\sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t \, \widehat{R}_g \, s_{gt}} \, (\, \overline{E} - E_A \,), \quad \forall \, f \, \in \, \mathcal{F}. \tag{10}$$

If $\widehat{R}_f = 1$ for all f, then allowances allocations are directly proportional to sales; unlike Rule II, however, here firms recognize that an increase in sales will directly increase the number of allowances received, irrespective of the amount of capacity.

In order for the above rules to be well-defined irrespective of whether the denominator is zero, we write

$$\alpha_f \, cap_f \;=\; \begin{cases} \alpha \, R_f \, cap_f & \text{for (8)} \\ \widehat{\alpha} \, \widehat{R}_f \displaystyle\sum_{t \in \mathcal{T}} H_t \, s_{ft} & \text{for (9)} \end{cases} \tag{11}$$

and

$$\alpha_f \sum_{t \in \mathcal{T}} H_t \, s_{ft} = \widehat{\alpha} \, \widehat{R}_f \sum_{t \in \mathcal{T}} H_t \, s_{ft} \quad \text{for (10)} \tag{12}$$

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 expressions in (11). In the rest of the paper, we focus on pure versions of rules I, II, and III.

The above descriptions of the contingent rules assume that investors know with certainty what allowances they will receive in advance of building a plant. In turn, this implies that they also know how much other capacity will be built (and, in the case of Rules II and III, how all capacity will be operated) since the fixed pool of allowances has to be shared among all plants. This assumption is consistent with the concept of a long-run equilibrium, in that if these profit expectations were not fulfilled then more or less entry would occur until expectations were consistent with realized profits, and those profits were driven to zero. We also assume that the generators receive their allowances according to the relevant rule at the same time they produce and sell power; this is a simplification of what actually occurs in that allocation commitments are usually made prior to that time.

2.4 Model solution

In the case of capacity-based Rules I and II, the model seeks a set of electricity sales $(s_{ft})_{(f,t)\in\mathcal{F}\times\mathcal{T}}$, capacities $\left(cap_f\right)_{f\in\mathcal{F}}$, 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 to the emission allocation rules I and II, respectively, the following conditions:

$$\begin{array}{lcll}
0 \leq s_{ft} & \perp & H_t\left[-\pi_t\left(\sum_{g\in\mathcal{F}} s_{gt}\right) + MC_{ft} + p_e E_f\right] + \mu_{ft} \geq 0, & \forall\,(f,t)\in\mathcal{F}\times\mathcal{T} \\
0 \leq \mu_{ft} & \perp & cap_f - s_{ft} \geq 0, & \forall\,(f,t)\in\mathcal{F}\times\mathcal{T} \\
0 \leq cap_f & \perp & -p_c - p_e\,\alpha_f + F_f - \sum_{t\in\mathcal{T}} \mu_{ft} \geq 0, & \forall\, f\in\mathcal{F} \\
0 \leq p_e & \perp & \overline{E} - e_{NP}(p_e) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\,E_g\,s_{gt} \geq 0 & \\
0 \leq p_c & \perp & \sum_{g\in\mathcal{F}} cap_g - \underline{CAP} \geq 0 & \\
(11) & \text{and} & \sum_{g\in\mathcal{F}} \alpha_g\,cap_g - (\,\overline{E} - E_A\,) = 0. &
\end{array} \tag{13}$$

For sales-based output Rule III, the first, third and last two conditions in the above problem are replaced by the analogous conditions in (4), (7) and (12). In particular, solutions to the model with Rule III need to satisfy the following conditions:

$$\begin{array}{lcll}
0 \leq s_{ft} & \perp & H_t\left[-\pi_t\left(\sum_{g\in\mathcal{F}} s_{gt}\right) + \mathrm{MC}_{ft} + p_e\,E_f - p_e\,\alpha_f\right] + \mu_{ft} \geq 0, & \forall\,(f,t)\in\mathcal{F}\times\mathcal{T} \\
0 \leq \mu_{ft} & \perp & cap_f - s_{ft} \geq 0, & \forall\,(f,t)\in\mathcal{F}\times\mathcal{T} \\
0 \leq cap_f & \perp & -p_c + F_f - \sum_{t\in\mathcal{T}} \mu_{ft} \geq 0, & \forall\, f\in\mathcal{F} \\
0 \leq p_e & \perp & \overline{E} - e_{NP}(p_e) - \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\,E_g\,s_{gt} \geq 0 & \\
0 \leq p_c & \perp & \sum_{g\in\mathcal{F}} cap_g - \underline{\mathrm{CAP}} \geq 0 & \\
(12) & \text{and} & \sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\,\alpha_g\,s_{gt} - \left(\overline{E} - E_A\right) = 0. &
\end{array} \tag{14}$$

3 Nonlinear Complementarity Formulations

To establish the existence of a solution to the model with the emission rules I, II and III, we derive equivalent formulations of (13) and (14) under these rules as standard nonlinear complementarity problems (NCPs).

3.1 Allowance allocation rule I

For the capacity-based rule I: $\alpha_f cap_f = \alpha R_f cap_f$ for all $f \in \mathcal{F}$, as given by (11). We now introduce a reformulation of (13) that replaces this rule by a complementarity condition. Specifically, we multiply the emission allowance balancing constraint $\sum_{f \in \mathcal{F}} \alpha_f cap_f = \overline{E} - E_A$ 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:

$$
\begin{array}{llll}
0 \leq s_{ft} & \perp & H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} s_{gt} \right) + MC_{ft} + p_e E_f \right] + \mu_{ft} \geq 0, & \forall \, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \leq \mu_{ft} & \perp & cap_f - s_{ft} \geq 0, & \forall \, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \leq cap_f & \perp & -p_c - \sigma \, R_f + F_f - \sum_{t \in \mathcal{T}} \mu_{ft} \geq 0, & \forall \, f \in \mathcal{F} \\
0 \leq p_e & \perp & \overline{E} - e_{NP}(p_e) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t \, E_g \, s_{gt} \geq 0 & \\
0 \leq p_c & \perp & \sum_{g \in \mathcal{F}} cap_g - \underline{CAP} \geq 0 & \\
0 \leq \sigma & \perp & \sigma \sum_{g \in \mathcal{F}} R_g \, cap_g - (\overline{E} - E_A) \, p_e \geq 0. &
\end{array}
\tag{15}
$$

Note the difference between (13) and (15): the former is a mixed NCP in the variable $(\mathbf{x}, \alpha)$, where

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

the latter is a standard NCP in the variable

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

To establish the relation between (15) and (13), we introduce the following mild condition on the sales of allowances to other sectors:

$$
\overline{E} > e_{NP}(0). \tag{16}
$$

This condition says that the number of allowances that the power sector can sell to nonpower sectors at any nonnegative price is strictly less than the total allowances allocated to power; this might, for example, be the result of regulatory limits.

We also impose a condition on firms' investment:

$$
\max \left(\underline{CAP}, \; \chi \max_{f \in \mathcal{F}} \left\{ \sum_{t \in \mathcal{T}} H_t \left[\pi_t(0) - MC_{ft} \right] - F_f \right\} \right) > 0, \tag{17}
$$

where χ is a unit converter to make the units consistent. Condition (17) implies that if $\underline{CAP} = 0$, then $\max_{f \in \mathcal{F}} \left\{ \sum_{t \in \mathcal{T}} H_t \left[\pi_t(0) - MC_{ft} \right] - F_f \right\} > 0$, which means that at least one firm will find it profitable to enter the market when quantity supplied in each period is zero. If no firm sells any power, then the power price will be expected to be 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.

The following result summarizes the connection between (15) and (13) with the emission rule I under the following two conditions (16) and (17).

Proposition 1. Under (16) and (17), if $(\mathbf{x}, \alpha)$ is a solution of (13) under the emission rule I, then

$$\sigma \equiv \frac{(\,\overline{E} - E_A\,)\, p_e}{\sum_{g \in \mathcal{F}} R_g\, cap_g} \tag{18}$$

is well defined and $\mathbf{x}^{\mathrm{I}}$ is a solution of (15). Conversely, if $\mathbf{x}^{\mathrm{I}}$ is a solution of (15), then

$$\alpha \equiv \frac{\overline{E} - E_A}{\sum_{g \in \mathcal{F}} R_g\, cap_g} \tag{19}$$

is well defined, and with $\alpha_f \equiv \alpha R_f$ for all $f \in \mathcal{F}$, $(\mathbf{x}, \alpha)$ is a solution of (13) under the emission rule I.

Proof. To prove the first assertion, let $(\mathbf{x}, \alpha)$ be as given. We first show that $cap_f > 0$ for some $f \in \mathcal{F}$. Suppose not, then $cap_f = s_{ft} = 0$ for all $(f,t) \in \mathcal{F} \times \mathcal{T}$, which implies $\underline{CAP} = 0$. We claim that $p_e = 0$. Indeed, if $p_e > 0$, then

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

which is a contradiction. From the first and third line in (13), we deduce, for all $f \in \mathcal{F}$,

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

or equivalently,

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

which contradicts (17). Therefore, the scalar σ in (18) is well defined; moreover, $\sigma = p_e \alpha$, yielding $\sigma R_f = p_e \alpha_f$. Consequently, (15) follows from (13). Conversely, let $\mathbf{x}^{\mathrm{I}}$ be a solution of (15). By the same argument as before, we deduce that $cap_f > 0$ for some $f \in \mathcal{F}$. Therefore, the scalar α in (19) is well defined; let $\alpha_f \equiv \alpha R_f$. We then have $\sum_{f \in \mathcal{F}} \alpha_f cap_f = \overline{E} - E_A$. Consequently, $(\mathbf{x}, \alpha)$ is a solution (13) under the emission rule I. □

It turns out that if $\underline{CAP} > 0$, then the NCP (15) 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

$$K \equiv \left\{ (\,\mathbf{s}, \mathbf{cap}\,) \geq 0 \,:\, \sum_{g \in \mathcal{F}} cap_g - \underline{CAP} \geq 0 \text{ and } cap_f - s_{ft} \geq 0,\ \forall\,(f,t) \in \mathcal{F} \times \mathcal{T} \right\}$$

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

$$\Phi^{\mathrm{I}}(\mathbf{s},\mathbf{cap},p_e) \;\equiv\; \left(\begin{array}{c} \left(H_t \left[-\pi_t \left(\displaystyle\sum_{g\in\mathcal{F}} s_{gt} \right) + MC_{ft} + p_e\, E_f \right] \right)_{(f,t)\in\mathcal{F}\times\mathcal{T}} \\ \left(F_f - \dfrac{(\,\overline{E} - E_A\,)\, p_e}{\displaystyle\sum_{g\in\mathcal{F}} R_g\, cap_g} R_f \right)_{f\in\mathcal{F}} \\ \overline{E} - \displaystyle\sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, s_{gt} - e_{NP}(p_e) \end{array} \right)$$

is a non-monotone map. Indeed, note that Φ^{I} is well defined on the set K^{I} because for every element $(\mathbf{s},\mathbf{cap}) \in K$, we must have $cap_f > 0$ 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 (15) under the identification (18) for σ. When $\underline{CAP} = 0$, the set

$$K \;=\; \prod_{f\in\mathcal{F}} \left\{ \left(\,(\,s_{ft}\,)_{t\in\mathcal{T}}, cap_f\,\right) \;\geq\; 0 \;:\; cap_f - s_{ft} \;\geq\; 0,\; \forall\, t \in \mathcal{T} \right\}$$

is the Cartesian product of separable sets. While the VI formulation is quite compact, one obvious advantage of the NCP (15) is that it applies to the case where $\underline{CAP} = 0$; in the latter case, the set K^{I} contains the origin where the function Φ^{I} fails to be well defined.

3.2 Allowance allocation rule II

The NCP formulation for (13) under the emission rule II is somewhat different. For one thing, it is unnecessary to have a change of variables in the formulation; in particular, we retain the variables α_f. The derivation in this subsection permits $\underline{CAP} = 0$. Specifically, consider the NCP:

$$\begin{array}{llll}
0 \,\leq\, s_{ft} & \perp & H_t \left[-\pi_t \left(\displaystyle\sum_{g\in\mathcal{F}} s_{gt} \right) + MC_{ft} + p_e E_f \right] + \mu_{ft} \geq 0, & \forall\,(f,t) \in \mathcal{F}\times\mathcal{T} \\
0 \,\leq\, \mu_{ft} & \perp & cap_f - s_{ft} \,\geq\, 0, & \forall\,(f,t) \in \mathcal{F}\times\mathcal{T} \\
0 \,\leq\, cap_f & \perp & -p_c - \alpha_f\, p_e + F_f - \displaystyle\sum_{t\in\mathcal{T}} \mu_{ft} \,\geq\, 0, & \forall\, f \in \mathcal{F} \\
0 \,\leq\, \alpha_f & \perp & \alpha_f\, cap_f - \widehat{\alpha}\, \widehat{R}_f \displaystyle\sum_{t\in\mathcal{T}} H_t\, s_{ft} \,\geq\, 0, & \forall\, f \in \mathcal{F} \\
0 \,\leq\, p_e & \perp & \overline{E} - e_{NP}(p_e) - \displaystyle\sum_{(g,t)\in\mathcal{F}\times\mathcal{T}} H_t\, E_g\, s_{gt} \,\geq\, 0 & \\
0 \,\leq\, p_c & \perp & \displaystyle\sum_{g\in\mathcal{F}} cap_g - \underline{CAP} \,\geq\, 0 & \\
0 \,\leq\, \widehat{\alpha} & \perp & \displaystyle\sum_{g\in\mathcal{F}} \alpha_g\, cap_g - (\,\overline{E} - E_A\,) \,\geq\, 0 &
\end{array} \tag{20}$$

in the variable $(\mathbf{x},\widehat{\alpha})$. The following result establishes the equivalence of the above NCP with (13) under the emission rule II.

Proposition 2. A pair $(\mathbf{x}, \widehat{\alpha})$ is a solution of (13) under the emission rule II if and only if it is a solution of (20).

Proof. The "only if" statement is obvious. Conversely, if $(\mathbf{x}, \alpha)$ is a solution of (20), it suffices to show that the following equalities hold:

$$\sum_{g \in \mathcal{F}} \alpha_g \, cap_g - (\,\overline{E} - E_A\,) \;=\; 0 \quad \text{and} \quad \alpha_f \, cap_f - \widehat{\alpha}\, \widehat{R}_f \sum_{t \in \mathcal{T}} H_t \, E_f \, s_{ft} \;=\; 0, \; \forall\, f \,\in\, \mathcal{F}.$$

Indeed if the first equality does not hold, then $\widehat{\alpha} = 0$ by complementarity, yielding $0 \leq \alpha_f \perp \alpha_f cap_f \geq 0$. In turn, this implies $\alpha_f cap_f = 0$ for all $f \in \mathcal{F}$; thus $\overline{E} - E_A = 0$, which contradicts the assumption that $\overline{E} > E_A$. Similarly, if $\alpha_f cap_f - \widehat{\alpha}\widehat{R}_f \sum_{t \in \mathcal{T}} H_t E_f s_{ft} > 0$ for some $f \in \mathcal{F}$, then $\alpha_f = 0$ by complementarity, which contradicts the inequality itself. $\square$

Similar to the VI $(K^{\mathrm{I}}, \Phi^{\mathrm{I}})$, if there exists a condition that can guarantee $cap_f > 0$ for all $f \in \mathcal{F}$, the NCP (20) is equivalent to the KKT conditions of the VI $(K^{\mathrm{II}}, \Phi^{\mathrm{II}})$, where $K^{\mathrm{II}} \equiv K^{\mathrm{I}} = K \times \Re_+$ and

$$\Phi^{\mathrm{II}}(\mathbf{s}, \mathbf{cap}, p_e) \;\equiv\; \left(\begin{array}{c} \left(H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} s_{gt} \right) + MC_f + p_e \, E_f \right] \right)_{(f,t) \in \mathcal{F} \times \mathcal{T}} \\[2ex] \left(F_f - \dfrac{(\,\overline{E} - E_A\,)\, p_e \, \widehat{R}_f \sum_{t \in \mathcal{T}} H_t \, s_{ft}}{cap_f \sum_{g \in \mathcal{F}} \widehat{R}_g \sum_{t \in \mathcal{T}} H_t \, s_{gt}} \right)_{f \in \mathcal{F}} \\[2ex] \overline{E} - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t \, E_g \, s_{gt} - e_{NP}(p_e) \end{array} \right).$$

3.3 Allowance allocation rule III

By the similar proof as that of Proposition 2, it can be shown that the a solution $(\mathbf{x}, \widehat{\alpha})$ of (14) under allowance allocation rule III is also a solution of the following NCP:

$$\begin{array}{llll}
0 \,\leq\, s_{ft} & \perp & H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} s_{gt} \right) + \mathrm{MC}_{ft} + p_e \, E_f - p_e \, \alpha_f \right] + \mu_{ft} \geq 0, & \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \,\leq\, \mu_{ft} & \perp & cap_f - s_{ft} \,\geq\, 0, & \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \\
0 \,\leq\, cap_f & \perp & -p_c + F_f - \sum_{t \in \mathcal{T}} \mu_{ft} \,\geq\, 0, & \forall\, f \in \mathcal{F} \\
0 \,\leq\, p_e & \perp & \overline{E} - e_{NP}(p_e) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t \, E_g \, s_{gt} \,\geq\, 0 & \\
0 \,\leq\, p_c & \perp & \sum_{g \in \mathcal{F}} cap_g - \underline{\mathrm{CAP}} \,\geq\, 0 & \\
0 \leq \alpha_f & \perp & \alpha_f - \widehat{\alpha}\, \widehat{R}_f \,\geq\, 0, & \forall\, f \in \mathcal{F} \\
0 \,\leq\, \widehat{\alpha} & \perp & \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t \, \alpha_g \, s_{gt} \,-\, (\overline{E} - E_A) \,\geq\, 0, &
\end{array} \tag{21}$$

Moreover, any solution of the NCP (21) must have $s_{ft} > 0$ for some $(f,t) \in \mathcal{F} \times \mathcal{T}$. Otherwise, suppose that some solution of NCP (21) has $s_{ft} = 0$ for all $(f,t) \in \mathcal{F} \times \mathcal{T}$. Then the last complementarity of NCP (21) becomes

$$0 \leq \widehat{\alpha} \perp -(\overline{E} - E_A) \geq 0,$$

which contradicts the condition $\overline{E} > E_A$.

Similar to the previous two allowance allocation rules, NCP (21) is equivalent to the KKT conditions of the VI $(K^{\mathrm{III}}, \Phi^{\mathrm{III}})$, where $K^{\mathrm{III}} \equiv K^{\mathrm{I}} = K \times \Re_+$ and

$$\Phi^{\mathrm{III}}(\mathbf{s}, \mathbf{cap}, p_e) \equiv \begin{pmatrix} \left(H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} s_{gt} \right) + MC_f + p_e \left(E_f - \dfrac{\widehat{R}_f \, (\overline{E} - E_A)}{\sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t \widehat{R}_g s_{gt}} \right) \right] \right)_{(g,t) \in \mathcal{F} \times \mathcal{T}} \\ (\, F_f \,)_{f \in \mathcal{F}} \\ \overline{E} - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t \, E_g \, s_{gt} - e_{NP}(p_e) \end{pmatrix}.$$

Since any solution of the NCP (21) must have $s_{ft} > 0$ for some $(f,t) \in \mathcal{F} \times \mathcal{T}$, the function $\Phi^{\mathrm{III}}(\mathbf{s}, \mathbf{cap}, p_e)$ is well defined.

4 Existence of Solutions

For the analysis in this section, we impose condition (16) and the following slight strengthening of (17):

$$\max_{f \in \mathcal{F}} \left\{ \sum_{t \in \mathcal{T}} H_t \left[\, \pi_t \, (0) - MC_{ft} \, \right] - F_f \right\} > 0. \tag{22}$$

This condition states that at least one firm would find investment profitable (fixed and variable cost less than revenue) if every generator produces no power. 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 (13) is nontrivial.

Proposition 3. *Under (16) and (22), any solution of the NCP (15) and (20) must have $s_{ft} > 0$ for some $(f,t) \in \mathcal{F} \times \mathcal{T}$.*

Proof. We prove the proposition only for (20). Assume for the sake of contradiction that some solution of this NCP has $s_{ft} = 0$ for all $(f,t) \in \mathcal{F} \times \mathcal{T}$. We claim that $p_e = 0$. Indeed, if $p_e > 0$, then

$$0 = \overline{E} - e_{NP}(p_e) \geq \overline{E} - e_{NP}(0) > 0,$$

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

$$\begin{aligned} 0 \quad &\leq \quad \sum_{t \in \mathcal{T}} \{ \, H_t \, [-\pi_t \, (0) + MC_{ft} \,] + \mu_{ft} \, \} \\ &\leq \quad \sum_{t \in \mathcal{T}} H_t \, [-\pi_t \, (0) + MC_{ft} \,] + F_f, \end{aligned}$$

which contradicts (22). □

We recall that the temporal price functions $\pi_t(\bullet)$ are strictly decreasing and the nonpower allowance demand function $e_{NP}(\bullet)$ is nonincreasing. The following is the main existence theorem for the model (13) with the emission rules I and II.

Theorem 4. Under conditions (16) and (22), the model (13) has a solution; under conditions (16), the model (14) has a solution.

We prove the above theorem via the three equivalent NCPs: (15) for emission rule I, (20) for emission rule II and (21) for emission rule III. In turn, the proofs for these three NCPs are quite similar. All 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 (20) 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 (15).

The existence proof for the NCP (21) is very also similar to the one for the NCP (20), except for an additional step concerning the boundedness of $\{p_e^k\}$, which we omit for brevity.

4.1 Proof for the NCP (20)

Toward the proof of solution existence to the NCP (20), we consider a perturbation of the function Φ^{II} in order to deal with the general case where some cap_f might be zero. Specifically, for each $\varepsilon > 0$, let

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

which is well defined on the set K^{II}. We first show that the VI $(K^{\text{II}}, \Phi^{\text{II}}_{\varepsilon})$ 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 $(\mathbf{s}^{\varepsilon,k}, \mathbf{cap}^{\varepsilon,k}, \boldsymbol{\mu}^{\varepsilon,k}, p_e^{\varepsilon,k}, p_c^{\varepsilon,k})$ be a tuple satisfying

$$0 \le s_{ft}^{\varepsilon,k} \;\perp\; H_t \left[-\pi_t \left(\sum_{g \in \mathcal{F}} s_{gt}^{\varepsilon,k} \right) + MC_{ft} + p_e^{\varepsilon,k} E_f \right] + \mu_{ft}^{\varepsilon,k} + \tau_k s_{ft}^{\varepsilon,k} \ge 0,$$

$$\forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \tag{23}$$

$$0 \le \mu_{ft}^{\varepsilon,k} \;\perp\; cap_f^{\varepsilon,k} - s_{ft}^{\varepsilon,k} + \tau_k\, \mu_{ft}^{\varepsilon,k} \;\ge\; 0, \quad \forall\, (f,t) \in \mathcal{F} \times \mathcal{T} \tag{24}$$

$$0 \le cap_f^{\varepsilon,k} \;\perp\; -p_c^{\varepsilon,k} + F_f - \frac{(\overline{E} - E_A) p_e^{\varepsilon,k} \widehat{R}_f \sum_{t \in \mathcal{T}} H_t s_{ft}^{\varepsilon,k}}{(cap_f^{\varepsilon,k} + \varepsilon) \left(\varepsilon + \sum_{g \in \mathcal{F}} \widehat{R}_g \sum_{t \in \mathcal{T}} H_t s_{gt}^{\varepsilon,k} \right)} - \sum_{t \in \mathcal{T}} \mu_{ft}^{\varepsilon,k} + \tau_k cap_f^{\varepsilon,k} \ge 0,$$

$$\forall\, f \in \mathcal{F} \tag{25}$$

$$0 \le p_e^{\varepsilon,k} \;\perp\; \overline{E} - e_{NP}(p_e^{\varepsilon,k}) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t\, E_g\; s_{gt}^{\varepsilon,k} + \tau_k\, p_e^{\varepsilon,k} \;\ge\; 0 \tag{26}$$

$$0 \le p_c^{\varepsilon,k} \;\perp\; \sum_{g \in \mathcal{F}} cap_g^{\varepsilon,k} - \underline{CAP} + \tau_k\, p_c^{\varepsilon,k} \;\ge\; 0. \tag{27}$$

We claim that under condition (16), the sequence $\{(\mathbf{s}^{\varepsilon,k}, \mathbf{cap}^{\varepsilon,k}, \boldsymbol{\mu}^{\varepsilon,k}, p_e^{\varepsilon,k}, p_c^{\varepsilon,k})\}$ is bounded. We show this in several steps: first the sequence $\{p_e^{\varepsilon,k}\}$; next the sequence $\{s_{ft}^{\varepsilon,k}\}$ for all $(f,t) \in \mathcal{F} \times \mathcal{T}$; then the sequence $\{cap_f^{\varepsilon,k}\}$ 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

$$\lim_{k (\in \kappa) \to \infty} p_e^{\varepsilon,k} \;=\; \infty. \tag{28}$$

Without loss of generality, we may assume that $p_e^{\varepsilon,k} > 0$ for all $k \in \kappa$. It follows by complementarity condition (26) that

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

For any $k \in \kappa$ such that $s_{f_0 t_0}^{\varepsilon,k} > 0$ for some pair (f_0, t_0), we have by complementarity condition (23)

$$H_{t_0} \left[-\pi_{t_0} \left(\sum_{g \in \mathcal{F}} s_{g t_0}^{\varepsilon,k} \right) + MC_{f_0 t_0} + p_e^{\varepsilon,k}\, E_{f_0} \right] + \mu_{f_0 t_0}^{\varepsilon,k} + \tau_k s_{f_0 t_0}^{\varepsilon,k} \;=\; 0,$$

which yields

$$p_e^{\varepsilon,k} \;\le\; E_{f_0}^{-1} \left[\, \pi_{t_0}(0) - MC_{f_0 t_0} \,\right]. \tag{29}$$

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

$$\begin{aligned} 0 &= \overline{E} - e_{NP}(p_e^{\varepsilon,k}) + \tau_k \, p_e^{\varepsilon,k} \\ &\geq \overline{E} - e_{NP}(0) \, > \, 0, \end{aligned}$$

which contradicts (16). Consequently, the bound (29) holds for all $k \in \kappa$, contradicting the limit (28).

Boundedness of $\{s_{ft}^{\varepsilon,k}\}$ 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 $\kappa \subset \{1, 2, \cdots, \}$,

$$\lim_{k (\in \kappa) \to \infty} s_{f_0 t_0}^{\varepsilon,k} \, = \, \infty. \tag{30}$$

Without loss of generality, we may assume that $s_{f_0 t_0}^{\varepsilon,k} > 0$ for all $k \in \kappa$. It follows by complementarity condition (23) that

$$\begin{aligned} 0 &= H_{t_0} \left[-\pi_{t_0} \left(\sum_{g \in \mathcal{F}} s_{g t_0}^{\varepsilon,k} \right) + MC_{f_0 t_0} + p_e^{\varepsilon,k} \, E_{f_0} \right] + \mu_{f_0 t_0}^{\varepsilon,k} + \tau_k \, s_{f_0 t_0}^{\varepsilon,k} \\ &\geq H_{t_0} \left[-\pi_{t_0} \, (0) + MC_{ft} + p_e^{\varepsilon,k} \, E_{f_0} \right] + \max(\, \mu_{f_0 t_0}^{\varepsilon,k}, \, \tau_k \, s_{f_0 t_0}^{\varepsilon,k} \,) \end{aligned}$$

which implies

$$\max(\, \mu_{f_0 t_0}^{\varepsilon,k}, \, \tau_k \, s_{f_0 t_0}^{\varepsilon,k} \,) \, \leq \, H_{t_0} \left[\pi_{t_0} \, (0) - MC_{ft} - p_e^{\varepsilon,k} \, E_{f_0} \right].$$

Since the right-hand side is bounded, by (30), it follows that

$$\lim_{k (\in \kappa) \to \infty} \tau_k \, = \, 0. \tag{31}$$

But this contradicts the inequality in (26): $\overline{E} - e_{NP}(p_e^{\varepsilon,k}) - \sum_{(g,t) \in \mathcal{F} \times \mathcal{T}} H_t E_g s_{gt}^{\varepsilon,k} + \tau_k p_e^{\varepsilon,k} \geq 0$. Therefore, $\{s_{ft}^{\varepsilon,k}\}$ is bounded for all $(f,t) \in \mathcal{F} \times \mathcal{T}$.

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

$$\lim_{k (\in \kappa) \to \infty} cap_{f_0}^{\varepsilon,k} \, = \, \infty. \tag{32}$$

Thus, by complementarity condition (27), we must have $p_c^{\varepsilon,k} = 0$ for all $k \in \kappa$ sufficiently large. Since $\{s_{f_0 t}^{\varepsilon,k}\}$ is bounded, we deduce $\mu_{f_0 t}^{\varepsilon,k} = 0$ for all $t \in \mathcal{T}$ and all $k \in \kappa$ sufficiently large. Without loss of generality, we may assume that $cap_{f_0}^{\varepsilon,k} > 0$ for all $k \in \kappa$. It follows by complementarity condition (25) that for all $k \in \kappa$ sufficiently large,

$$F_{f_0} - \frac{(\overline{E} - E_A) \, p_e^{\varepsilon,k} \, \widehat{R}_{f_0} \displaystyle\sum_{t \in \mathcal{T}} H_t \, s_{f_0 t}^{\varepsilon,k}}{(\, cap_{f_0}^{\varepsilon,k} + \varepsilon \,) \left(\varepsilon + \displaystyle\sum_{g \in \mathcal{F}} \widehat{R}_g \sum_{t \in \mathcal{T}} H_t \, s_{gt}^{\varepsilon,k} \right)} + \tau_k \, cap_{f_0}^{\varepsilon,k} \, = \, 0,$$

which implies that

$$F_{f_0} \;\le\; \frac{(\overline{E} - E_A)\, p_e^{\varepsilon,k}\, \widehat{R}_{f_0} \sum_{t\in\mathcal{T}} H_t\, s_{f_0 t}^{\varepsilon,k}}{(\, cap_{f_0}^{\varepsilon,k} + \varepsilon\,) \left(\varepsilon + \sum_{g\in\mathcal{F}} \widehat{R}_g \sum_{t\in\mathcal{T}} H_t\, s_{gt}^{\varepsilon,k} \right)}.$$

The limit (32) implies that the right-hand side tends to zero as $k(\in \kappa) \to \infty$, which is a contradiction. Hence $\{cap_f^{\varepsilon,k}\}$ 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 $\kappa \subset \{1, 2, \cdots, \}$,

$$\lim_{k(\in\kappa)\to\infty} \mu_{f_0 t_0}^{\varepsilon,k} \;=\; \infty. \tag{33}$$

Without loss of generality we may assume that $\mu_{f_0 t_0}^{\varepsilon,k} > 0$ for all $k \in \kappa$. By complementarity condition (24), we deduce $cap_{f_0}^{\varepsilon,k} - s_{f_0 t_0}^{\varepsilon,k} + \tau_k \mu_{f_0 t_0}^{\varepsilon,k} = 0$. Since $\{(cap_{f_0}^{\varepsilon,k}, s_{f_0 t_0}^{\varepsilon,k})\}$ is bounded, (33) implies that

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

Since

$$\begin{aligned} \mu_{f_0 t_0}^{\varepsilon,k} \;\le\; \sum_{t\in\mathcal{T}} \mu_{f_0 t}^{\varepsilon,k} \;\; &\le\;\; -p_c^{\varepsilon,k} + F_{f_0} - \frac{(\overline{E} - E_A) p_e^{\varepsilon,k}\, \widehat{R}_{f_0} \sum_{t\in\mathcal{T}} H_t\, s_{f_0 t}^{\varepsilon,k}}{(cap_{f_0}^{\varepsilon,k} + \varepsilon) \left(\varepsilon + \sum_{g\in\mathcal{F}} \widehat{R}_g \sum_{t\in\mathcal{T}} H_t\, s_{gt}^{\varepsilon,k} \right)} + \tau_k cap_{f_0}^{\varepsilon,k} \\ &\le\;\; F_{f_0} + \tau_k\, cap_{f_0}^{\varepsilon,k} \end{aligned}$$

and $\tau_k cap_{f_0}^{\varepsilon,k} \to 0$ as $k(\in\kappa) \to \infty$, we obtain a contradiction to (33).

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 $\{(\mathbf{s}^{\varepsilon,k}, \mathbf{cap}^{\varepsilon,k}, \boldsymbol{\mu}^{\varepsilon,k}, p_e^{\varepsilon,k}, p_c^{\varepsilon,k})\}$ under the condition (16). 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 $(\mathbf{s}^\varepsilon, \mathbf{cap}^\varepsilon, p_e^\varepsilon)$ be one such solution. For each $\varepsilon > 0$, there exists $(\mu_{ft}^\varepsilon, p_c^\varepsilon)$ such that

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

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

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

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

$$\frac{\sum_{t \in \mathcal{T}} H_t \, s_{ft}^{\varepsilon_k}}{(\, cap_f^{\varepsilon_k} + \varepsilon_k \,)} \leq \frac{\sum_{t \in \mathcal{T}} H_t \, cap_f^{\varepsilon_k}}{(\, cap_f^{\varepsilon_k} + \varepsilon_k \,)} < \sum_{t \in \mathcal{T}} H_t$$

it follows that the sequence $\left\{ \frac{\sum_{t \in \mathcal{T}} H_t \, s_{ft}^{\varepsilon_k}}{(\, cap_f^{\varepsilon_k} + \varepsilon_k \,)} \right\}$ must have at least one accumulation point. Without loss of generality, we may assume that this sequence converges to a limit, say $\gamma_f \geq 0$. Note that

$$\gamma_f = \frac{\sum_{t \in \mathcal{T}} H_t \, \widehat{s}_{ft}}{\widehat{cap}_f}, \quad \text{if the denominator is positive.}$$

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

$$\alpha_f \equiv \frac{(\overline{E} - E_A) \, \widehat{R}_f \, \gamma_f}{\sum_{g \in \mathcal{F}} \widehat{R}_g \sum_{t \in \mathcal{T}} H_t \, \widehat{s}_{gt}} \, \widehat{p}_e.$$

It is easy to show that the following complementarity holds:

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

With

$$\widehat{\alpha} \equiv \frac{(\overline{E} - E_A) \, \widehat{p}_e}{\sum_{g \in \mathcal{F}} \widehat{R}_g \sum_{t \in \mathcal{T}} H_t \, \widehat{s}_{gt}},$$

it is easy to see that all conditions in (20) are satisfied.

5 Illustrative Application

We now illustrate the results and sometimes counterintuitive insights that can be obtained from our proposed models about the market performance of different allowance allocation systems. Contingent

Rules I, II, and III are compared to auctioning/grandfathering of allowances. The effect of the presence of capacity markets and minimum output constraints for baseload plants upon the conclusions is also examined. As mentioned before, the NCPs (15), (20), and (21) 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: $e_{NP}(p_e) = 0$;
• Generator types: $i = 1$ (coal steam), 2 (natural gas-fired combined cycle), and 3 (natural gas-fired combustion turbine);
• Marginal costs: $\text{MC}_{1t} = 20$ \$/MWh, $\text{MC}_{2t} = 40$ \$/MWh, and $\text{MC}_{3t} = 80$ \$/MWh, for all $t \in \mathcal{T}$;
• 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; and
• Total capacity requirement: $\underline{\text{CAP}} = 11{,}000$ MW, if $\underline{\text{CAP}} > 0$.

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 every period. A somewhat high number of periods are chosen so that changes in capacity mix will result in a relatively smooth response in terms of the number of periods that each type of generator is the marginal source. 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 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.

The technology types chosen represent a range of fossil-fueled conventional generators now under construction in various markets. Their construction and fuel costs are representative of prevailing values in 2005. Emission rates are typical for those types of plants. In the computations, the NCPs (15), (20), and (21) and the linear complementarity problem (34) below are solved by the *PATH* solver available on the *NEOS* server (http://neos.mcs.anl.gov/neos/solvers/index.html).

5.1 Solution performance measures

We introduce the following system performance measures for comparing the solutions:

• *Generation cost* (M\$/yr), total investment and fuel costs: $\displaystyle\sum_{f \in \mathcal{F}} \left(F_f cap_f + \sum_{t \in \mathcal{T}} H_t MC_{ft} s_f \right)$;

• *Social cost* (M\$/yr), defined as the decrease in social surplus relative to a base case with no emissions constraints, where social surplus is defined as PS + CS + AS. PS is the equilibrium producer surplus, equal to the sum over all firms of the objectives in (1); CS is the equilibrium consumer surplus: $\displaystyle\sum_{t \in \mathcal{T}} H_t \left(\int_0^{d_t(p_t)} \pi_t(x)dx - d_t(p_t)p_t \right)$; and AS is the auction (or grandfathering) surplus, equal to the the revenue received by the government if it auctioned those allowances (or, equivalently, the economic rent accruing to the original owners of grandfathered allowances). Social surplus can also be shown to equal the value of consumption (integral of all demand curves) minus generation cost. We deliberately exclude external environmental costs from this performance measure because the emphasis of our analysis is on comparing solutions with the same emissions but different social costs;

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

• *Capacity factor*, the ratio of annual generation to potential generation for each plant type f: $\dfrac{\sum_{t\in\mathcal{T}} H_t s_{ft}}{\sum_{t\in\mathcal{T}} H_t cap_f}$.

We do not report producer surplus PS separately, because it is by assumption zero in the free entry solutions.

For the purpose of assessing the impact of different allowance allocation schemes, we focus on the increase in generation cost, social cost, and consumer payments relative to the (emission) unconstrained solution obtained using the LCP in the next subsection (Subsection 5.2). In the case in which a capacity constraint is imposed, these values are 2049 M$/yr, 20911 M$/yr, and 2049 M$/yr, respectively (Table 1). For ease of comparison, those increases are expressed as a percentage of the generation cost of the unconstrained solution. The three percentage increases are calculated, respectively, as (Table 1):

• relative generation cost increase: 100% (generation cost - 2049 M$/yr)/2049 M$/yr;

• relative social cost increase: 100% [20,911 M$/yr - (PS + CS + AS)]/2049 M$/yr; and

• relative consumer payments increase: 100% (consumer payments - 2049 M$/yr)/2049 M$/yr.

For the market simulations in which there is no capacity market, the percent changes are instead calculated relative to the costs and social surplus in an emissions unconstrained solution that also lacks a capacity market. In that case, the base amounts of generation cost, social cost, and consumer payments are instead 1893, 21020, and 1893 M$/yr, respectively (Table 1). Finally, for the market simulations in which minimum run constraints are imposed on coal plants, the base amounts are 2278, 20628, and 2278 M$/yr respectively (last row, Table 1).

Imposing a CO_2 constraint results in a social cost in the form of a loss of social surplus relative to the unconstrained value of 20,911 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 identical to the change in generation cost. Of course, the CO_2 constraint also provides unquantified social benefits in the form of environmental improvement; however, as mentioned, we focus on the differences in social cost between scenarios with the same emissions. Another unquantified benefit is the additional reliability that might be provided by having a capacity requirement of 1100 MW in the capacity market solutions; although those solutions might be more expensive than the solutions without such a requirement, there may be compensating reliability improvements which we do not quantify. Thus, the most meaningful comparisons will be of scenarios in Table 1 with the same CO_2 limit (either 20 or 40 MTons/yr) and the same capacity constraint (either 11,000 MW or absent).

5.2 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 $\overline{E} = \infty$, the constraint (5) 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:

$$
\begin{aligned}
0 &\le s_{ft} \quad \perp \quad H_t\left[-\pi_t\left(\sum_{g\in\mathcal{F}} s_{gt}\right) + MC_{ft}\right] + \mu_{ft} \ge 0, && \forall (f,t)\in\mathcal{F}\times\mathcal{T}\\
0 &\le \mu_{ft} \quad \perp \quad cap_f - s_{ft} \ge 0, && \forall (f,t)\in\mathcal{F}\times\mathcal{T}\\
0 &\le cap_f \quad \perp \quad -p_c + F_f - \sum_{t\in\mathcal{T}} \mu_{ft} \ge 0, && \forall f\in\mathcal{F}\\
0 &\le p_c \quad \perp \quad \sum_{g\in\mathcal{F}} cap_g - \underline{CAP} \ge 0.
\end{aligned}
\tag{34}
$$

The solution is shown in Table 1 as the Base Run row after Run 18. 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, equal to the total of energy and capacity revenues divided by total quantity demand, is 46.1 $/MWh. Of that amount, capacity payments make up 27%. The sum of consumer surplus and producer surplus is 20,911 M$/yr.

5.3 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, II, and III 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 100%, 50%, and 0% allowances granted to capacity or sales on the basis of Rules I, II, or III; i.e., $E_A/\overline{E} = 0,\ 0.5,\ 1$, respectively. We take R_f in (8) and $\widehat{R}_f$ in (9) to be both equal to E_f/E_1. Finally, $\widehat{R}_f$ in (10) is set equal to 1, so that allowances are simply allocated in proportion to sales under Rule III.

In the presence of a capacity market, runs 1–6 show the results for the capacity-based potential emissions Rule I, runs 7–12 are results for the capacity-based actual emissions Rule II, and runs 13-18 are the output-based sales Rule III. Runs 19-36 are the same runs, but for the case of no capacity market. The last three rows in Table 1 (Runs 37-39) are Rule I results with a capacity market for the case of min-run constraints for coal plants.

First we discuss the capacity market runs 1-18. The cases with 0% allowances freely distributed are the same for all three contingent allocation rules because, of course, that no allowances are allocated to new investment or sales. In the case where all allowances are instead allocated to new plants by one of the capacity-based rules (Rules I, II; Runs 1, 4, 7, 10), we see that investment is greatly distorted and costs are much higher than if no allowances are granted freely to new entry. More coal capacity is built, and then this expensive capacity is used inefficiently, with much lower capacity factors than in the case where new investment receives no free allowances.

The potential emissions Rule I has the worst distortion, because it is possible for new generators to receive free allowances even if they generate no power. Under that rule, 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 under Rule I, 95% of the combustion turbine capacity is never used, and is built just to collect free allowances. Unexpectedly, the distortion is relatively worse in the mild (40 Mton/yr) CO_2 limit case on a $/ton of reductions basis, making the total cost of compliance almost as high as for the 20 Mton/yr case under Rule I. Hence it should not be assumed that the risk of distortion is less for less severe CO_2 limits.

On the other hand, basing free allocation of allowances upon sales (Rule III) rather than capacity

(Rules I, II) yields only mild distortions in costs and capacity mix under the capacity market assumption. There are slight (at most 8%) increases in combined cycle plant (and accompanying decreases in coal capacity) relative to the 0% contingently allocated solution, and the social cost of meeting the carbon constraint is slightly higher. This shift in mix toward cleaner capacity is made necessary because lower energy prices (due to the output subsidies implicit in the allowances allocation rule) have increased demand that must be met while still complying with the cap (as in [7]). Thus, in a single area market (neither imports nor exports), it appears that allocating allowances freely based on sales is much less distorting than allocating allowances to new capacity construction.

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 19–36), 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 review here. Instead we merely consider the interaction of capacity and emissions markets. The main effect we observe in comparing Runs 19-36 with Runs 1-18 is in the case of Rule II, the capacity-based 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 Rule I (capacity-based potential emission rule), the costs and generation mixes in the case where 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 so much overinvestment that the capacity constraint no longer binds.

Thus, as Smeers and Ehrenmann [23] indicate, it is possible for the overinvestment bias arising from giving away allowances to capacity to have a compensating benefit of increasing capacity reserves. This could help correct market failures that result in too little capacity in energy markets. In other runs of our model (not shown), we have examined the effect of imposing a tight price cap ($100/MWh) upon energy prices in each period in the absence of a capacity market. The price cap-constrained equilibrium is calculating using the method described in [11]. If no free allowances are granted to new plants by Rule I , the result is significantly less capacity (400-500 MW decrease) compared to Runs 21 and 24 (0% allowances contingently allocated, no price cap, no energy market). But if instead 100% of allowances are allocated by Rule II under the tight energy price cap, the decrease in capacity is only 100 MW under the 20 MT cap; so in that case, the inefficient allocation rule has largely erased the capacity disincentive due to energy price caps. However, the 500 MW gap in capacity due to the price cap is not erased by Rule II in the 40 MT case, so this "second best" effect does not always occur.

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 over the course of a day in the way they are in the Table 1 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 37-39 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) are diminished somewhat, but remain large. In that case, the cost of complying with the 20 Mton/yr limit is more than doubled compared to the equilibrium yielded if none of the allowances are contingently allocated.

5.4 Detailed results for Rule II: Capacity-based actual emissions rule

To take a more detailed look at the effect of contingent allocation of allowances, we solved a set of problems with varying percentages of allowances contingently allocated by Rule II, assuming the presence of a capacity market. 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). Along the horizontal axis of each figure, we decrease the fraction of emissions that are freely granted by Rule II from 100% to 0%. 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; consequently generators pay nothing, on net, for their emissions. On the other hand, at the right-hand extreme, new capacity has to pay for 100% of their emissions, either buying them from a government auction or purchasing them from owners of grandfathered allowances.

Two sets of results are shown for each emissions cap: Figures 2 and 3 show the effect of contingent allocation upon three categories of costs as well as 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 100%, 50% and 0% allowances allocated by Rule II are already reported in Runs 7–12 in Table 1.)

Note that consumer payments are in excess of generation investment and fuel costs if less than 100% of the allowances are contingently allocated (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 auctioned (or grandfathered) 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 government who auctions the allowances (or the original owners of the grandfathered allowances, if instead those allowances are grandfathered). Figure 3 shows that in the case with 0% freely allocated allowances, the 40 Mton/yr limit has a very small social cost: 19 M$/yr, or less than 1% of total investment and operating cost. However, the social cost of the stricter 20 Mton/yr limit is much larger, 395 M$/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 of meeting the limit are inflated if some or all allowances are freely given to new entry under the contingent Rule II. The distortions, as measured by generation or social cost, are mild until the fraction of contingently allocated allowances exceeds 80%, and then increase rapidly as that fraction increases further. In the extreme case of 100% of allowances being freely allocated, the costs of meeting 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, if no allowances are given away, 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, under the loose CO_2 constraint, the contingent allocation of allowances to new entry has inflated the cost of meeting the CO_2 constraint by more than an order of magnitude.

Further, 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 granted by Rule II increases, the price of emissions allowances also 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) at which natural gas plants are cheaper to run, on the margin, than coal plants, once the opportunity cost of allowances is factored into the cost. The dispatch order is then reversed compared to the case where new capacity receives no free allowances, 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 capacity-based 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 100% contingent allocation of allowances than under 0% contingent allocation (see also Table 1). This is partially consistent with the conjecture of the Netherlands Bureau for Economic Policy Analysis [15] that much of the value of freely-granted allowances is passed back to consumers in the long run. But not all that value is returned; much of it is instead eaten up by the dead weight loss of investment distortions. If the resulting economic rent associated with auctioned/grandfathered allowances could be returned to consumers either via tax reductions or other mechanisms, then consumers would generally be much better off in the case where no allowances are granted freely to new capacity than if most allowances are allocated by capacity-based contingent allocation rules.

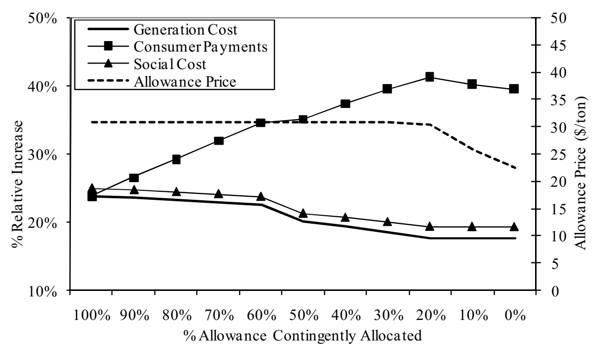

Figure 2: Cost and price comparison, under capacity- based allocation rule II (20 Mton limit)

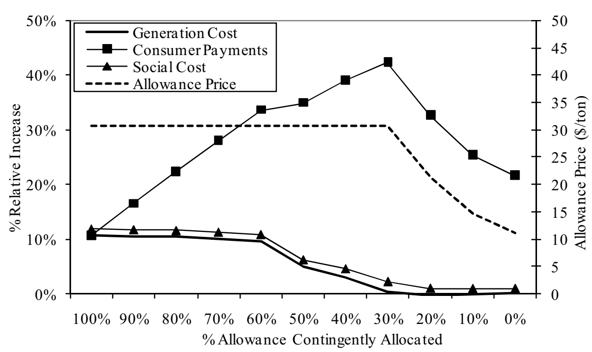

Figure 3: Cost and price comparison, under capacity- based allocation rule II (40 Mton limit)

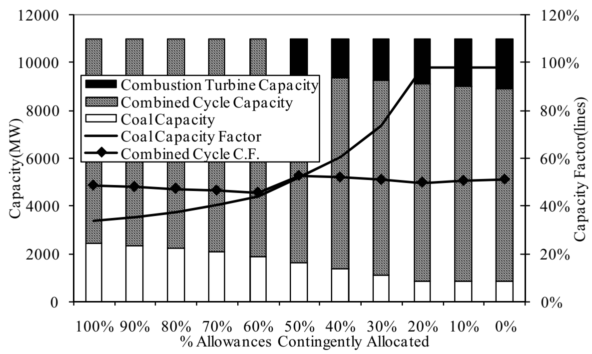

Figure 4: Capacity and capacity factor comparison, under capacity- based allocation rule II (20 Mton limit)

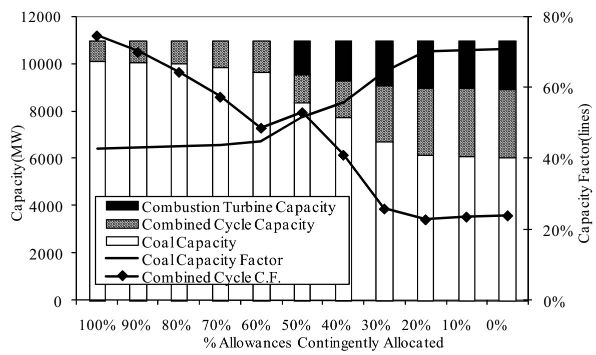

Figure 5: Capacity and capacity factor comparison, under capacity- based allocation rule II (40 Mton limit)

6 Conclusion

We have presented complementarity problem formulations for long-run markets for electric power and emissions permits for the purpose of analyzing the efficiency of alternative systems for allocating emissions allowances. Existence of solutions was proven under mild conditions. A simple numerical example illustrates the potential for investment distortion arising from allocation rules that give allowances to new capacity; under the assumptions made, less distortion occurs if instead allowances are allocated in proportion to energy sales, and the least distortion occurs if allowances are not freely distributed by contingent allocation rules. Model runs show that the absence or presence of capacity markets, price caps on energy prices, and minimum run constraints upon baseload generator output can significantly affect these conclusions.

Future work should address formulation of more realistic models including, for instance, transmission or carbon sequestration alternatives; parameterization based on actual markets; non-steady state dynamics in terms of technology, cost, and demand changes; and representation of interlinked markets in which different markets are subject to different rules or no rules at all, as is presently the case in the European Union and the Regional Greenhouse Gas Initiative, respectively. An important issue is possible interactions of market power and allocation rules. Large generation firms have the potential to manipulate prices; it is possible that allowing firms to strategically recognize the endogeneity of allocations under some rules will lead to even greater distortions. The models formulated here readily lend themselves to consideration of market power in the form of Cournot competition among generators both for inputs (e.g., allowances) and outputs (energy and capacity) e.g., [20].

References

[1] Ahman, M., K. Holmgren. 2006. New entrant allocation in the Nordic energy sectors: Incentives and options in the EU ETS. *Climate Policy* **6** 423–440.

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

[3] Burtraw, D., K. Palmer, R. Bharvirkar, 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] Demailly, D., 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] Dissou, Y. 2005. Cost-effectiveness of the performance standard system to reduce CO_2 emissions in Canada. *Resource and Energy Economics* **27** 187–207.

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

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

[8] Fischer, C., A.K. Fox. 2007. Output-based allocation of emissions permits for mitigating tax and trade interactions. *Land Economics* **83** 575–599.

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

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

[11] Hobbs, B.F., J.S. Pang. 2007. Nash-Cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints, *Operations Research* **55** 113–127.

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

[13] Matthes, F., V. Graichen, 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.

[14] Metzler, C.B., B.F. Hobbs, 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.

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

[16] Neuhoff, K., 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, N. Walker. 2006. Implications of announced phase II national allocation plans for the EU ETS, *Climate Policy* **6** 411–422.

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

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

[19] Palmer, K., D. Burtraw, D. Kahn. 2006. Simple rules for targeting CO_2 allowance allocations to compensate firms. *Climate Policy* **6** 477–493.

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

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

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

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

[24] Sterner, T., A. Muller. 2008. Output and abatement effects of allocation readjustment in permit trade. *Climatic Change* **86** 33–49.

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

Table 1. Summary of Model Results

Capa-city Market (MW)	Min Output Level	Allow-ance Allocation Rule	Run	CO_2 Limit (Mton/yr)	%Cont-ingent Alloca-tion	% Increase Relative to Base Run			Capacity(MW)			% Capacity Factor			Quantity-weighted mean p_t ($/MWh)	p_e ($/ton)	p_c ($/MW/yr)
						Gen. Cost	Social Cost	Consumer Payments	Coal	Comb. Cycle	Comb. Turbine	Coal	Comb. Cycle	Comb. Turbine			
11000	None	I. Capacity-Based, Potential Emission Rule	1	20	100%	28.7%	31.6%	28.7%	3088	5540	5361	27.6%	73.2%	0.3%	61.11	34.35	0
			2		50%	16.9%	19.4%	31.3%	863	7747	2390	97.8%	52.6%	0.9%	62.15	29.54	14617
			3		0%	17.6%	19.3%	39.5%	852	8084	2064	97.8%	51.1%	0.3%	65.65	22.45	50000
			4	40	100%	26.5%	29.2%	26.5%	8094	559	12155	53.9%	99.2%	0.1%	59.97	32.46	0
			5		50%	-0.2%	2.5%	29.8%	6803	1729	2468	64.0%	32.0%	1.0%	61.52	30.77	8462
			6		0%	0.1%	0.9%	21.6%	6076	2871	2053	71.1%	23.9%	0.5%	56.70	11.01	50000
		II. Capacity-Based, Actual Emission Rule	7	20	100%	23.8%	25.0%	23.8%	2459	8541	0	33.7%	48.6%	N/A	58.11	30.77	29115
			8		50%	20.1%	21.2%	35.1%	1600	7889	1511	51.8%	52.7%	0.0%	63.40	30.77	50000
			9		0%	Same results as Runs 3, 15, 39											
			10	40	100%	10.6%	11.8%	10.6%	10140	860	0	42.8%	74.6%	N/A	51.93	30.77	4593
			11		50%	5.0%	6.2%	35.0%	8359	1211	1430	51.9%	53.0%	0.0%	63.37	30.77	50000
			12		0%	Same results as Runs 6, 18											
		III. Output-Based Rule	13	20	100%	18.8%	19.4%	18.8%	821	8212	1967	98.0%	51.3%	0.3%	55.32	22.45	50000
			14		50%	18.2%	19.3%	29.2%	837	8148	2015	97.9%	51.2%	0.3%	60.46	22.45	50000
			15		0%	Same results as Runs 3, 9, 39											
			16	40	100%	1.1%	1.0%	1.1%	5958	3084	1958	72.0%	24.9%	0.5%	46.66	11.01	50000
			17		50%	0.6%	0.9%	11.4%	6017	2978	2005	71.6%	24.4%	0.5%	51.65	11.01	50000
			18		0%	Same results as Runs 6, 12											
		Base Run		No CO_2 Limit		2049	20911	2049	7329	1628	2042	65.10%	17.65%	0.59%	46.13	0.00	50000
None	None	I. Capacity-Based, Potential Emission Rule	19	20	100%	Same as results as Run 1											N/A
			20		50%	22.8%	23.8%	39.1%	1976	6557	290	43.2%	61.9%	5.6%	61.02	30.77	
			21		0%	18.7%	20.6%	42.4%	886	7601	0	97.7%	53.3%	N/A	62.61	22.45	
			22	40	100%	Same as results as Run 4											
			23		50%	5.1%	5.8%	38.2%	8038	553	301	54.3%	99.3%	5.0%	60.62	31.43	
			24		0%	-0.1%	0.9%	23.2%	6160	2365	0	70.7%	25.7%	N/A	53.68	11.01	
		II. Capacity-Based, Actual Emission Rule	25	20	100%	25.3%	24.8%	25.3%	2226	6851	0	37.8%	60.2%	N/A	54.58	30.77	N/A
			26		50%	21.6%	22.5%	37.9%	1606	7109	0	53.4%	57.3%	N/A	60.45	30.77	
			27		0%	Same results as Runs 21,33											
			28	40	100%	17.1%	15.9%	17.1%	9752	807	0	44.5%	79.5%	N/A	50.77	30.77	
			29		50%	4.2%	4.9%	36.7%	7859	889	0	55.6%	63.4%	47.3%	59.89	30.77	
			30		0%	Same results as Runs 24, 36											
		III. Output-Based Rule	31	20	100%	20.3%	20.8%	20.3%	855	7730	0	97.8%	53.5%	N/A	52.31	22.45	N/A
			32		50%	19.5%	20.7%	31.3%	871	7666	0	97.8%	53.4%	N/A	57.43	22.45	
			33		0%	Same results as Runs 21, 27											
			34	40	100%	1.3%	1.0%	1.3%	6040	2580	0	71.6%	26.7%	N/A	43.66	11.01	
			35		50%	0.6%	0.9%	12.2%	6100	2473	0	71.1%	26.2%	N/A	48.64	11.01	
			36		0%	Same results as Runs 24, 30											
		Base Run		No CO_2 Limit		1893	21020	1893	7329	1232	0	65.10%	19.79%	N/A	43.10	0.00	N/A
11000	Coal Only	I. Cap-Based Potential	37	20	100%	13.4%	13.5%	13.4%	1441	7018	5811	58.5%	57.8%	0.6%	59.75	30.77	0
			38		50%	5.1%	5.0%	18.1%	Same results as Run 2								
			39		0%	5.8%	4.9%	25.5%	Same results as Runs 3, 9, 15								
		Base Run		No CO_2 Limit		2278	20628	2278	2287	6671	2042	0.9%	0.4%	0.0%	51.68	0.00	50000