Allocating Emissions Permits in Cap-and-trade Programs: Theory and Evidence *

Preliminary.

Please do not cite without permission

Comments very welcome

December 2009

Meredith Fowlie†

Abstract

The allocation of emissions permits in "cap-and-trade" programs is an increasingly contentious policy design issue. Recent theoretical work has characterized the efficiency and distributional implications of alternative approaches to allocating these permits in detail. This paper addresses the empirical question: Is theory borne out in practice? I develop a simple analytical model to capture the essential theoretical relationships between permit allocation design choices and firm-level production decisions. Data gathered from a multi-state emissions trading program are used to analyze these relationships empirically. Results suggest that firms do account for explicit environmental compliance costs (i.e. the costs of holding permits to offset emissions) in their short-run supply decisions. The data provide somewhat weaker evidence that firms are also responding to the less salient production subsidy that is implicitly conferred by dynamic permit allocation updating regimes.

* I thank seminar participants at Environmental Defense, the University of California, Berkeley,the University of Michigan, the University of Toronto, and Yale University for helpful comments and discussion. I gratefully acknowledge support from the National Science Foundation through grant SES-0922401, the UC Energy Institute and the Center for State, Local, and Urban Policy at the University of Michigan. Erik Johnson provided excellent research assistance.

†Department of Agricultural and Resource Economics, University of California Berkeley, 301 Giannini Hall, Berkeley CA, 94720. email: fowlie@berkeley.edu

1 Introduction

Billions of dollars worth of tradable emissions permits are allocated each year to U.S. industrial producers regulated under emissions "cap-and-trade" programs.[1] In theory, how these permits are allocated can have significant implications for who will bear the costs and how efficiently the mandated emissions reductions will be achieved. Permit allocation has thus emerged as one of the more contentious issues in permit market design.

Regulatory agencies have been allocating tradable emissions permits under the auspices of local, regional, and nationwide cap-and-trade programs for over a decade. Over this time period, theoretical analyses of the efficiency and distributional implications of permit market design choices has grown increasingly sophisticated. Accumulated theory and experience notwithstanding, we know relatively little about how permit allocation design affects firm decision-making in real world settings. Some core assumptions underlying policy simulation models are largely untested. This paper brings evidence to bear on a first-order empirical question: are firms responding to permit allocation incentives as theory predicts?

Traditionally, policy makers have chosen between two general approaches to allocating emissions permits: auctioning and grandfathering. Under an auction regime, emissions permits are sold to the highest bidder. In contrast, "grandfathered" permits are freely distributed to regulated sources based on pre-determined, firm-specific characteristics. In the absence of other market failures, this choice between grandfathering and auctioning should have no bearing on permit market efficiency in the short-run (Montgomery, 1974).

Many economists favor auctioning on the grounds that revenues can be used to offset distortionary factor taxes (Crampton and Kerr, 2002; Goulder *et al.*,1999).[2] However, in practice, policy makers have routinely chosen to forego auction revenues in favor of handing permits out for free

[1] The value of permits allocated to emitting facilities under the NOx Budget Program and the Acid Rain Program each year is roughly $1.4 B and $4.5B, respectively. The U.S. Environmental Protection Agency (EPA) has estimated that the value of allowances allocated annually under proposed Federal climate legislation would exceed $200 billion (US EPA, 2008) .

[2] Other efficiency-related arguments in favor of auctioning pertain to the mitigation of pre-existing regulatory distortions and distributional concerns. For example, Dinan and Rogers (2002) and Parry (2004)**?** emphasize the potential distributional implications of the allocation design choice, demonstrating that high income individuals are likely to gain more from freely allocated allowances than are low income individuals,

to regulated entities.[3] The ability to make concessions to adversely impacted and politically powerful stakeholders via grandfathering has arguably been as important a factor in the widespread adoption of emissions trading programs as the promise of cost minimization and gains from trade.

More recently, a third design alternative has emerged. Under a "contingent allocation" regime, updating rules established ex ante are used to determine how a firm's permit allocations will be periodically updated over the course of the trading program. Allocation updating is typically based on a firm's production choices (such as output levels or fuel inputs). The incentives created by contingent allocation rules are quite different from those associated with grandfathering or auctioning because updating creates an incentive to increase whatever activity determines future emissions permit allocations.

In a theoretical, "first-best" setting, it is straightforward to demonstrate that periodically updating firms' future permit allocations based on present production choices will undermine the efficiency of permit market outcomes because the implicit subsidy conferred by allocation updating encourages firms to increase output to economically inefficient levels. (Bohringer and Lange, 2005; Sterner and Muller, 2008).[4] However, contingent updating can welfare dominate other permit allocation approaches when there are additional, pre-existing distortions to contend with. For example, the theory literature has explored how allocation updating can be used reduce inefficiencies resulting from the exercise of market power (Fischer, 2003; Gerbasch and Requate, 2004; Neuhoff, Martinez, and Sato, 2006), tax interaction effects (Fischer and Fox, 2007), and emissions leakage (Bernard *et al.,* 2007; Quirion and Demailly, 2006). Allocation updating can also been discussed as a used to reduce impacts on consumer prices and mitigate effects on sectoral adjustment to address political concerns about the incidence of a cap-and-trade programs (Jensen and Rasmussen, 2000).

Political support for contingent allocation updating is increasing. Industry groups endorse it as a "common sense way to promote efficiency, fairness, and environmental protection".[5] Policy

[3] A majority of permits are distributed freely to regulated entities in Southern California's RECLAIM program, the European Union's Emissions Trading Program, the nation-wide Acid Rain Program, and the regional NOx Budget Trading Program.

[4] Here, first best refers to a regulatory environment in which the only market distortion or imperfection is the environmental externality that the emissions regulation is designed to internalize.

[5] "Advantages of Allocating Emissions Credits Based on Efficiency." The Clean Energy Group. May 15, 2009.

experts concede that it may offer the most pragmatic approach to mitigating the adverse effects of environmental regulation on domestic industry competitiveness when emissions regulations are incomplete.[6] Federal climate legislation passed in June 2009 by the House of Representatives includes provisions for allocation updating as a means of compensating trade exposed emitters for compliance costs incurred.[7] In, legislators are considering allocation updating as a means of mitigating impacts on consumer prices and reducing emissions leakage to unregulated entities (Bushnell and Chen, 2009).[8]

Can permit allocation design be effectively used to achieve these kinds of policy objectives? This depends in large part on whether firms respond to permit market incentives as standard theory predicts. In policy debates, stakeholders have begun to question the extent to which the implicit subsidy conferred by updating will be factored into firms' "real world" production decisions.[9] Others contend that these subsidy effects will be mitigated by pre-existing regulatory and market constraints (NCEP, 2008). Finally, researchers investigating private-sector decision-making in other contexts have documented gain-loss asymmetries and heuristic behaviors. If these behavioral factors influence environmental compliance decisions, firms may perceive and respond to permit allocation incentives in unexpected ways(see, for example, Duxbury and Summers, 2004; Hirshliefer, 2001).

This study makes three contributions to both the academic literature and the ongoing policy discourse. In the first part of the paper, a simple analytical model is used to illustrate the essential short-run implications of the permit allocation design. In an attempt to clarify the terms of the policy debate surrounding this issue, these theoretical implications have been over-simplified to the point of misrepresentation. For example, in an attempt to "clear up misperceptions, common

http://www.thecleanenergygroup.com/lsgbriefings.asp>In

[6]See, for example, the testimony of Richard Morgenstern of Resources for the Future. *Competitiveness and Climate Policy: Avoiding Leakage of Jobs and Emissions*: *Hearings before the committee on Energy and Commerce U.S. House of Representatives.* March 18, 2009

[7]The justification for contingent updating in this context rests on the concern that firms would divert new investments and production to manufacturing facilities located in countries without commensurate regulations. Contingent allocation updating, intended as a stop-gap measure, compensates firms for the compliance costs incurred so as to mitigate adverse competitiveness impacts.

[8]The design recommendations of both the California Public Utilities Commission and theWCI include the minimization of the impacts of carbon regulations on consumers as a prominent objective of the allocation process. In this context, allocation updating has emerged as a politically palatable option (Bushnell and Chen, 2009).

[9]See, for example, RGGI, 2004.

among many stakeholders, about how allocation decisions do and do not affect the way an emissions trading program works in practice", an influential report asserts that "[a]llocation affects the distribution of benefits and burdens among firms and industry sectors; it does not change program results or overall costs (NCEP, 2007)." In fact, this conventional wisdom does not hold when future permit allocations are contingent upon current production choices. The simple model presented here intuitively demonstrates how permit allocation design can significantly affect firms' compliance choices (and thus aggregate social costs).

Second, the paper offers some of the first empirical evidence of how contingent allocation updating is working in practice. Previous attempts to study the effects of contingent permit allocation updating, and revenue recycling more generally, have been unable to convincingly separate the effects of the implicit subsidy from the overall effect of the environmental regulation.[10] I exploit an unusual policy setting in which the implicit subsidies conferred by a regional emissions trading program vary systematically across producers. More importantly, much of this cross-sectional and inter-temporal variation is exogenous- in an econometric sense- to firms' short-run production decisions. I present a general model of the underlying data generating process to motivate a more descriptive analysis of firms' observed esponse to permit allocation incentives. I then derive a tractable reduced form that can be implemented empirically using a discrete choice framework. This facilitates a more structural analysis of how permit allocation design decisions have affected firms' short run compliance decisions. Estimation results provide strong evidence that firms account for explicit compliance costs (i.e. the costs of holding permits to offset emissions) in their short-run production decisions. The data provide somewhat weaker evidence that firms are also accounting for the implicit subsidy, although there appears to be substantial variation in the extent to which different plants incorporate this positive incentive in their supply decisions. Attempts to explain this variation in terms of observable plant characteristics meet with only limited success;

[10] Sterner and Isaksson (2006) were the first to empirically investigate the effects of revenue recycling in the context of market-based emissions regulation. They analyze a Swedish program in which emissions charges are refunded to polluting firms in proportion to output. Because rebates do not vary across firms or across time, the authors cannot separate the effect of the tax fro the effect of the recycled revenues. In a more recent paper, Sjim *et al.* (2006) encounter similar difficulties in their analysis of how sequentially grandfathered permits in the EU ETS impacts electricity market outcomes. The lack of clarity surrounding how current production decisions will influence future permit allocations in the European Union's Emissions Trading System complicates their analysis of how this implicit updating has affected firm decision making.

I provide weak evidence that larger plants are more responsive to these positive incentives.

Finally, the empirical results provide insights into how permit market incentives affect firms' short-run production decisions more generally. An emerging literature examines the relationships between permit price series and wholesale electricity price series (Bunn and Fezzi, 2007; Fell, 2009; Sijm et al., 2006 ; Zachmann and von Hirschhausen, 2008). However, given the complexity of interactions between permit markets and wholesale electricity markets, it has been difficult to deduce firm-level behaviors from wholesale market price dynamics. This paper examines firm-level responses to changing permit market conditions in unprecedented detail.

The paper proceeds as follows. The following section develops a simple theoretical model that is used to intuitively demonstrate the first order implications of alternative permit allocation designs. Section 3 introduces the NOx Budget Program. Section 4 discusses the data that were collected from this program. Section 5 augments the simple framework presented in the first section so as to reflect some important features of the underlying data generating process and introduces the empirical strategy. Section 6 summarizethe estimation results. Section 7 concludes.

2 Allocating emissions permits: Theory

The analytical framework introduced in this section serves two purposes. The model is first used to demonstrate the most essential partial equilibrium implications of the permit allocation design choice, and then to motivate the empirical analysis. The model is intentionally simple. Many of the insitutional details and market imperfections captured by models found elsewhere in the literature (such as pre-existing tax distortions, the exercise of market power, or incomplete regulation) have been stripped away. Eliminating some of the complexities of real policy settings helps to highlight the the most basic trade-offs between static production efficiency, static allocative efficiency, and distributional concerns.

The analysis will focus on short run relationships exclusively. To the extent that allocation updating is seen as a way to smooth the transition to auctioning regimes, these short run relationships will be very important. Also, a clear characterization of short run interactions is an essential first step towards understanding longer-run outcomes and implications.

2.1 A simple partial equilibrium framework

Production of a homogeneous good generates harmful pollution. Industry production in time t is denoted Q_t. Let q_{it} denote the quantity produced by firm i in time t. Producers are characterized by increasing marginal cost technologies; $C_i(q_{it})$ and $c_i q_{it}$ denote the unit-specific total cost and marginal cost functions, respectively. Efficient factor markets are assumed. Marginal operating costs thus reflect the true opportunity cost of allocating inputs to production in this industry. Emissions rates e_i are constant per unit of output but vary across firms. Demand is characterized by an affine inverse demand function $P_t = a - bQ_t$.

To keep the model transparent and tractable, only two price taking firms are represented. A more general model with $N > 2$ is easily formulated but more difficult to intuitively interpret. The two firms are indexed c (denoting the relatively "clean" producer) and d (denoting the relatively "dirty" producer). For the purpose of this example, I assume emissions rates are negatively correlated with operating costs: ; $e_c < e_d$,. $c_c > c_d$.[11]

Industry emissions are regulated under a cap-and-trade program; aggregate emissions in period t cannot exceed an exogenously determined cap E_t. The time path of permitted emissions (*i.e.* $E_1, E_2, ...$) is set by the regulator ex ante. To comply with the program, firms must offset uncontrolled emissions with permits. These permits are tradable in an emissions permit market. There are no spatial or temporal restrictions on permit trading. I assume that firms acts as price takers in both the permit and product markets

This short run analysis conditions on existing production technology and operating characteristics; emissions rates and operating costs are exogenously determined and fixed. Emissions reductions can thus be achieved in two ways: through increasing the share of the market served by the relatively clean producer or reducing the quantity consumed.

In existing and planned cap-and-trade programs, permits are allocated via auctioning, grandfathering, symmetric "output-based" updating, or asymmetric updating. Stylized representations of all four approaches are considered below.

[11] This assumption finds empirical support. Unit-level fuel operating costs and NOx emissions rates are negatively correlated in the data analyzed in the subsequent section. However, there are certainly examples of low-emitting facilities with relatively low operating costs, and high-emitting facilities with relatively high operating costs.

2.2 The benchmark case

Outcomes under alternative allocation rules will be compared against a "first best" benchmark that maximizes total economic surplus $S(Q)$ subject to technology operating constraints and the constraint that aggregate emissions do not exceed the exogenously set cap:

$$\begin{aligned} \max_{q_{ct},q_{dt}} \quad : \quad & S(Q_t) = \int_0^{Q_t} P(Q_t)dQ_t - C_c(q_{ct}) \; - C_d(q_{dt}) \\ s.t. \;\; e_c q_{ct} + e_d q_{dt} \;\; = \;\; & E_t \\ q_{ct} + q_{dt} \;\; = \;\; & Q_t. \end{aligned} \tag{1}$$

Note that the welfare measure $S(Q_t)$ reflects the utility associated with total consumption less production costs but does not capture the benefits associated with emissions reductions. Because aggregate emissions are held constant at E across all scenarios, changes in $S(Q_t)$ across scenarios will reflect changes in absolute welfare vis a vis this benchmark.

The first order conditions for this maximization problem imply:

$$P(Q_t^*) - c_i q_{it}^* - \tau_t e_i = 0, \qquad i = c, d. \tag{2}$$

where τ_t is the shadow value of the emissions constraint at time t. The * superscript denotes values that maximize economic surplus subject to the constraint.

Rearranging these first order conditions (and omitting the t subscripts for expositional clarity) yields:

$$\frac{P - c_d q_d^*}{P - c_c q_c^*} = \frac{e_d}{e_c}. \tag{3}$$

Figure 1 helps to illustrate this result. The downward sloping line, representing the emissions constraint, connects all allocations of production across the two firms that exactly satisfy the emissions cap. The slope of this line is $\frac{e_d}{e_c}$. The economic surplus function $S(Q)$ is also projected into this space. The level sets of the surplus function appear as concentric iso-surplus curves. The slope of an iso-surplus curve measures the rate at which production at the clean firm can be substituted for production at the dirty firm while holding total economic surplus constant. The

socially optimal allocation of production occurs at the point where the emissions constraint is just tangent to an iso-surplus curve. All other points that exactly satisfy the emissions constraint are associated with lower levels of economic surplus.

Two efficiency properties of this equilibrium are worth highlighting:

Property 1 : Marginal abatement costs (measured in terms of foregone profits per unit of emissions reduction) are set equal across producers:

$$\frac{P - c_d q_d^*}{e_d} = \frac{P - c_c q_c^*}{e_c}. \tag{4}$$

This assures that abatement activities have been efficiently allocated among producers. Given production level Q^* , the cost of meeting the emissions constraint E is minimized.

Property 2 : Emissions abatement activities are allocated efficiently across producers and consumers:

$$\frac{\partial S^*}{\partial E} = \frac{c_c q_c^* - c_d q_d^*}{e_d - e_c}. \tag{5}$$

Intuitively, the derivative of the welfare function with respect to the emissions constraint captures the marginal cost of reducing emissions via conservation measures on the demand side (i.e. through a reduction in consumption). The marginal abatement cost on the supply side is the cost of reallocating production from the low cost, high emitting producer to the high cost, low emitting producer so as to incrementally reduce emissions. [5] implies that an optimal balance is struck between the two short run abatement options. Appendix 1 works through this result in more detail.

Returning to Figure 1, the broken line connects all points associated with an aggregate production level Q^*. Production allocations on the emissions constraint lying strictly above (below) the optimal outcome are associated with less (more) consumption than is consistent with compliance constrained economic surplus maximization, thus requiring higher (lower) levels of supply-side abatement to achieve compliance.

2.3 Grandfathering and auctioning regimes

I now consider a perfectly competitive industry subject to a market-based emissions cap-and-trade program. Let A_{it} represents the firm's permit allocation in period t. Under grandfathering, the number of permits the firm receives (free of charge) from the regulator each period is determined at the outset of the program. Under auctioning, $A_{it} = 0 \; \forall \; i$. Under either scenario, firms' future permit allocations are independent of their production decisions going forward.

Let τ_t represent the permit price (an endogenously determined parameter). The firm's compliance costs (i.e. the cost of holding permits to offset uncontrolled emissions) are $\tau_t(A_{it} - e_i q_{it})$. The profit maximization problem faced by price taking firm i in time period t is thus:

$$\max: \; \pi_{it} = P_t q_{it} - C(q_{it}) + \tau_t(A_{it} - e_i q_{it}), \qquad i = c, d. \tag{6}$$

Assuming price-taking behavior in both the permit and product markets, the first order conditions for this profit maximization problem are given by [2]. Thus, the efficiency properties [4] and [5] are satisfied under both grandfathering and auctioning.

2.4 Output-based updating

Under a contingent updating regime, the total quantity of emissions permits to be allocated in each period, and the rules specifying how firms' production decisions will determine future permit allocations, are determined at the outset of the program. An output-based updating regime defines a firm's permit allocation in period t as a function of the firms' production decisions in the preceeding period or periods. To simplify the analysis, I consider the simplest of output-based allocation rules: a firm's permit allocation in period $t+1$ is determined by its market share in period t:

$$A_{it+1} = \frac{E_{t+1}}{Q_t} q_{it} \equiv s_t q_{it}. \tag{7}$$

The size of the subsidy conferred by output-based updating will depend on the total number of permits allocated in the future period E_{t+1}, how the firm discounts future revenue streams $\delta_i(t)$, the future permit price τ_{t+1}, and total industry production Q_t.Contingent-updating thus adds an

additional argument to the firm's profit function.:

$$\max: \ \pi_{it} = P_{it}q_{it} - C(q_{it}) - \tau_t e_i q_{it} + \delta_i(1)\ \tau_{t+1} s_t q_{it}. \tag{8}$$

Some additional assumptions further simplify the analysis. Unrestricted banking and borrowing of permits, rational expectations, and zero arbitrage together imply that permit prices are constant in present value terms. If all firms discount future period gains at the market rate and take total sector output Q_t as given, the implicit subsidy per unit of production in period t simplifies to τs_t.[12] Omitting t subscripts for simplicity, the first order conditions for profit maximization under output-based updating are:

$$P(Q') - c_i q_i' - \tau(e_i + s) = 0, \quad i = c, d \tag{9}$$

The subscript ' denotes equilibrium outcomes under non-contingent permit updating. Rearranging [9] yields:

$$\frac{P - c_d q_d'}{P - c_c q_c'} = \frac{e_d - s}{e_c - s}.$$

This equilibrium outcome $\{\ q_c',\ q_d'\}$ lies strictly above the optimal outcome on the emissions constraint line and is associated with a level of total economic surplus that is strictly less than $S(Q^*)$ (see figure 1). Intuitively, the implicit production subsidy increases each firm's willingness to supply vis a vis grandfathering or auctioning. At higher levels of production, more supply-side abatement is required to comply with the emissions cap. Consequently, the market share of the relatively clean producer increases relative to the first best benchmark.

Appendix 2 demonstrates that the equilibrium permit price will be unambiguously higher than τ^*, reflecting higher supply-side marginal abatement costs. Consumption levels exceed that associated with optimal compliance, so there is a transfer of surplus from producers to consumers vis a vis grandfathering.[13]

[12] Alternatively, firms could take into account how their own production decisions affect aggregate production levels (and thus the size of the implicit subsidy). This would reduce the perceived production subsidy by $\frac{\delta_t \tau_t A_{t+1} q_t}{Q_t^2}$. Intuitively, if the firm incrementally increases production in time t, it decreases the number of permits allocated in time $t+1$ per share of output Q_t, although it is now entitled to an additional share.

[13] Making similar distributional comparisons with auctioning is complicated by the fact that consumer and pro-

2.5 Asymmetric contingent allocation

In practice, the implicit production subsidy introduced by contingent updating is often asymmetric. For political and practical reasons, the s parameter often varies across units based on technology type, fuel efficiency, or other observable operating characteristics. To accomodate asymmetric updating, the firm's objective function is modified slightly to allow the implicit production subsidy to vary across firms. Omitting the t subscripts for simplicity, the first order conditions for profit maximization are:

$$P(Q'') - c_i q_i'' - \tau(e_i + s_i) = 0, \quad i = c, d \tag{10}$$

I assume that the updating parameters are defined such that the total number of permits allocated through updating does exceed the total cap. This implies that the average updating parameter cannot exceed the average emissions rate. In this asymmetric case, the equilibrium permit price is no longer equal to the economic cost of redispatching production activity in order to incrementally reduce emissions. If the implicit production subsidy favors the relatively dirty (clean) firm, the permit price must rise above (fall below) the true marginal abatement cost in order to counteract this asymmetry in compliance incentives. The equilibrium permit price is:

$$\tau'' = \frac{c_c q_c - c_d q_d}{e_d - e_c + (s_c - s_d)}. \tag{11}$$

The efficiency and distributional implications of asymmetric updating vis a vis first best will depend on the ratio of the updating parameters s_c and s_d. If the implicit subsidy per unit of pollution is exactly equal across firms (i.e. implying that $\frac{s_d}{s_c} = \frac{e_d}{e_c}$), the first best quantities and product price are achieved (see Appendix 3). If the subsidy per unit of emissions is larger for the relatively clean firm (as is the case with output based updating), the outcome will occur at a point on the emissions constraint above the optimal outcome (implying too little conservation on the demand side and too much abatement on the supply side). Conversely, if the subsidy per unit of emissions is larger for the more polluting firm, consumer prices will increase and consumption levels will fall relative to first best. A larger share of the mandated emissions reductions will be

ducer surplus under auctioning will depend on how auction revenues are allocated.

achieved on the demand-side.

2.6 Motivating the empirical exercise

Conditional on the assumptions outlined above, contingent allocation updating will distort outcomes away from the efficient short-run equilibrium in a first best setting (except in the very special case where $\frac{s_i}{s_j} = \frac{e_i}{e_j}, \forall\ i \neq, j$). Abatement efforts will be inefficiently allocated across producers and consumers and across firms with different production technologies and cost stuctures.[14] Much of the theory literature is devoted to extending this kind of analytical exercise to more complicated, second-best settings. In cases where the implicit subsidy can be used to mitigate one or more pre-existing distortions or imperfections (such as the exercise of market power in the product market, or incomplete emissions regulation) contingent allocation updating can theoretically welfare dominate grandfathering and auctioning.

Underlying this burgeoning literature is the assumption that compliance cost minimizing firms in cap-and-trade programs accurately and equally account for all permit market incentives in their supply decisions However, there are several reasons why this assumption might not hold in practice. First, the behavioral finance literature offers evidence to suggest that firms may focus on information that is more readily accessible and easy to understand at the expense of information that is more opaque or that requires more resources to process (Hirshleifer, 2001; Sarin and Weber, 1993). It is arguably much easier for plant managers to translate permit prices into compliance costs per unit of production, versus the expected future subsidy per unit of current production. Particularly when allocation updating rules are convoluted or confusing.

Second, researchers have found evidence of gain loss asymmetry (whereby agents place more emphasis on minimizing losses versus maximizing gains) in private sector decision-making (Fiegenbaum, 1990; Gabel and Sinclair-Desgagne, 2000). Studies have also documented assymmetric cost transmission, whereby firms pass operating cost increases through to customers at a different rate than they pass any cost reductions). For instance, a recent study offer evidence of asym-

[14] General equilibrium models calibrated to specific policy contexts have been used to quantitatively estimate the potential magnitude of these distortions. In several instances, inefficiencies induced by allocation updating have been found to be economically significant. See, for example, Burtraw *et al.* (2005), Jensen and Rasmussen (2000), Neuhoff *et al.* (2005). (Jensen & Rasmussen 2000)(**?**)(**?**)(Quirion & Demailly 2008)

metric compliance cost past through in the EU Emissions Trading Scheme (Zachmann and von Hirschhausen, 2008). To the extent that these behavioral elements affect firms' environmental compliance decisions, managers may discount- or ignore- the implicit production subsidy.

Finally, if there is any uncertainty regarding how the cap-and-trade regulation will be implemented or modified in the future, managers may discount the implicit production subsidy to reflect this regulatory risk.

In sum, whereas it is standard to assume that the explicit compliance cost and implicity production subsidy are weighted equally in firms' production decisions, this may not be the case in practice.The obligation to hold permits to offset uncontrolled emissions may affect short run compliance decisions differently than the implicit subsidy conferred by permit allocation updating. If firms discount- or ignore- the implicit subsidy conferred by contingent allocation updating, permit allocation incentives will not have the intended effects on firms' short-run production decisions, or market outcomes. In what follows, this assumption is evaluated using detailed data from a regional emissions trading program.

3 Empirical application: The NOx Budget Program

In 1998, the U.S. Environmental Protection Agency (EPA) determined that 23 eastern states were contributing significantly to ozone non-attainment problems. These states were issued "NOx budgets" and required to design and implement regulations that would reduce seasonal NOx emissions to budget levels. Although states had flexibility in choosing their compliance strategies, they were invited to meet their compliance obligations by joining an EPA-administered cap-and-trade program. All states accepted the invitation.

States in the NOx Budget Program (NBP) are required to accept program design features outlined in a model rule that was issued by the EPA. These features include permit trading protocols and emissions reporting standards. For instance, throughout the program, a NOx permit authorizes the holder to emit one ton of NOx during "ozone season" (i.e. May to September).[15]

[15] Compliance is only required in the spring and summer when average temperatures rise and NOx emissions contribute to smog formation.

At the end of each season, all regulated source must hold sufficient permits to offset ozone season emissions.[16] There are no spatial trading restrictions in the NBP; permits are freely traded among all participating sources in all participating states.

The model rule also required standardization of intertemporal trading restrictions across participating states. Emitters cannot borrow against future allocations. Emissions in year t must be offset using permits of vintage t or earlier. Permits can be banked, although the use of banked permits is subject to a "progressive flow control" (PFC) constraint designed to discourage the excessive use of banked permits in a particular year.[17]

3.1 Permit allocation in the NBP

In the process of designing the NBP model rule, the US EPA commissioned an ex ante analysis of permit allocation design alternatives. A detailed simulation model of the electricity sector was used to evaluate various allocation regimes, including grandfathering, output-based allocation updating, and fuel input-based allocation updating (US EPA, 1999)[18]. Simulation results indicated that the permit allocation design choice would appreciably affect market outcomes. Simulated retail electricity prices were 3.4 percent lower under updating versus grandfathering (representing a transfer of $1.25 billion from producers to consumers). Supply-side abatement costs were projected to be 18 percent higher under updating versus grandfathering (largely due to an increased share of electricity supply provided by relatively clean- and relatively more costly- natural gas units). It was also anticipated that emissions leakage to neighboring, unregulated states would be reduced by allocation updating; simulated electricity production in regulated jurisdictions was 10 percent higher under updating as compared to grandfathering.

[16] If a facility's emissions exceed its permit allocation, the facility must purchase additional NOx permits in the permit market. Compliance has been nearly perfect over the duration of the program; the few cases of non-compliance have been attributed to accounting errors.

[17] By law, if the number of permits in the region-wide bank prior to ozone season exceeds 10 percent of the total (i.e. program-wide) cap for that season, a non-linear discount factor is applied. The PFC ratio is computed as 10 percent of the seasonal cap divided by the size of the bank. This ratio defines the fraction of banked permits that can be used to offset a ton of permits that season. The remaining permits can be used to offset only a half ton.The discount factor is applied at the facility level. For example, if a single firm holds 100 permits and the ratio in year t is defined to be 0.5, that firm can use 50 banked permits to offset emissions in year t on a one for one basis.

[18] Because over 90 percent of emissions regulated under this program come from electricity producers, EPA analysis focused exclusively on the electricity sector.

Whereas many important program design features were defined at the federal level, states were ultimately given broad flexibility with regards to permit allocation design. The EPA recommended allocation updating based on heat inputs, but states were free to deviate from this recommendation.[19] Several states chose to pursue alternative approaches; permit allocation rules vary significantly in terms of overall regime choice (i.e. grandfathering, or contingent updating), the frequency of allocations, and the basis for distributing the allowances.

4 Data

I use data from the first four years of the NOx Budget Program (i.e. 2003-2006) and focus exclusively on electricity producers serving restructured wholesale electricity markets in the Eastern United States (i.e. the New York, New England , and Mid-Atlantic or "PJM" markets).[20] This section briefly summarizes the data.

State-level permit allocation regimes

Permit allocation design parameters were collected from state-level implementing agencies. All states have documented the specific algorithms and equations used to determine facility-specific permit allocations in detail. Agencies were contacted directly when implementation details were unclear or could not be found in the public record.

Table 1 reports state-level NOx budgets (which were pre-determined and do not change over the study period) and information regarding state-specific permit allocation design choices. Whereas smaller states chose grandfathering (due in part to the management resources required to administer a more complex permit allocation updating process), a majority of states chose some form of contingent allocation updating based on either output or fuel inputs.

[19] Fuel-based updating was chosen over output-based updating primarily because, historically, emissions regulations had been defined in terms of mass emissions per unit of heat input.

[20] Electricity generating facilities (EGUs) comprise 87 percent of the emissions sources and over 90 percent of the NOx emissions regulated under the NBP (EPA, 2007). EGUs regulated under the NBP operate in a variety if electricity market environments. Whereas units in the sample supply restructured wholesale electricity markets, other facilities in the program are rate-regulated producers serving vertically integrated, economically regulated electricity markets, while other units are owned and operated by public entities and operate on a non-profit basis. Production at rate-regulated plants and public entities are more centrally coordinated and influenced by an array of economic, regulatory, and institutional factors. I choose to focus on restructured electricity markets because EGUs in these markets are more likely to have short-run objectives consistent with profit maximization.

Unit-level operations and attributes

Hourly, unit-level emissions, heat input, and output data were obtained from the US EPA Continuous Emission Monitoring System (CEMS).[21] Continuous emission monitoring and reporting requirements for EGUs regulated under EPA's Acid Rain Program (ARP) and/or the NOx Budget Program require monitoring of hourly sulfur dioxide mass emissions, carbon dioxide mass emissions, NOx emission rates, heat input, and other operating conditions.[22]

Table 2 summarizes some important unit-level operating characteristics by permit allocation regime. I exclude nuclear, hydro, and renewable fuels because these zero emitting producers are categorically excluded from the NOx budget program. Small producers (i.e. less than 10 or 15 MW) are also excluded from the NBP. Finally, I exclude any units for which wholesale electricity generation is not the primary production activity (such as self generating units). This leaves 610 electricity generating units in the cleaned data set.

Two operating attributes that are particularly relevant to this analysis are the NOx emissions rate (i.e. pounds of NOx emitted per MWh electricity produced) and heat rate (i.e.btus of fuel burned per kWh of electricity produced).[23] It is fairly standard in the empirical literature to treat these unit-specific performance parameters as immutable features of the production technology. However, emissions rates and heat rates can be affected by operating decisions made by the plant manager, including the choice of fuel characteristics, utilization rates, and combustion tuning. Purely exogenous factors (such as ambient temperature) can also play a role.

To construct unit-specific summary measures of these operating characteristics, separate seasonal regression equations are estimated for each unit. This estimation exercise, described in more detail in Appendix 4, obtains unit-specific, season-specific point estimates of emission

[21] Coal units report hourly gross and net unit load (in MWe). Combined-cycle units are required the sources to report the sum of all loads associated with all cycles (steam and electric) on a consistent basis. This sometimes involves converting steam load to an equivalent electrical load or vice-versa.

[22] Only those units that participate in both the NBP and the the Acid Rain Program are required to report hourly operations year round. Units regulated under the NBP only need not report outside of ozone season. Hourly data are also not reported when plants are taken off-line for scheduled outtages or maintenance. In total, the analysis sample includes 15.9 million hourly observations.

[23] A unit's heat rate measures the efficiency with which the unit transforms fuel into electricity. The lower the heat rate, the more fuel efficient the generator.

rates and heat rates under average operating conditions.[24] Capacity-weighted summaries of these estimates are presented in Table 2. The support of the distributions overlap considerably across allocation regimes, which facilitates an empirical comparison of short-run production decisions made by similar units facing different permit allocation incentives.

Emissions permit prices

NOx permits are actively traded in a liquid permit market.[25] Brokers facilitate the majority of arms length transactions, administer escrow accounts, and provide market analysis. A variety of transaction structures exist in the market, including forward settlements, calls, puts, composite structures such as straddles, and vintage swaps.

Daily permit price data were purchased from Evolution Markets LLC. Table 3 reports average spot NOx permit prices by vintage (in nominal dollars per ton). NOx permit prices are falling over the sample period, largely due to abatement costs that proved to be lower than anticipated, and lower than expected temperatures in the early years of the program.[26] This table also helps to illustrate the effect of the progressive flow control (PFC) constraint on permit prices. As early as 2003, permit market participants correctly anticipated that the PFC constraint would start to bind in 2005. This explains the large vintage 2004/2005 spread in 2003 and 2004.[27].

Compliance costs and production incentives

To estimate the cost of purchasing permits to offset the emissions associated with generating a MWh of electricity, each unit's NOx emissions rate is multiplied by the NOx permit price. Table 4 summarizes these unit-level cost estimates. On average, explicit compliance costs amount to a 7 percent increase in total variable (i.e. fuel, operating and maintenance) costs.[28] However, among units with particularly high emissions rates, this increase can exceed 40 percent.

[24] Specifications that allow rates to vary across years are also estimated.

[25] In 2007, the volume of "economically significant" immediate settlement trades (i.e. trades between versus within firms) reached 247,000 tons (EPA 2008).

[26] Evolution Markets LLC provides informative monthly analyses of the NOx Budget Program permit market.

[27] In years when the PFC constraint binds, banked permits trade at a considerable discount. The PFC ratio was 0.25 and 0.27 in 2005 and 2006, respectively. In both years, permits were used to offset emissions at a discounted rate (4,168 and 1,950 permits in 2005 and 2006, respectively). In March of 2005, the EPA released its new Clean Air Interstate Rule (CAIR), intended to subsume the NBP in 2009. CAIR eliminated progressive flow control.

[28] Unit-specific estimates of variable fuel operating costs are obtained by multiplying the unit-level heat rate (see above) by the corresponding fuel price. Estimates of unit-level variable, non-fuel operating and maintenance costs (not including environmental compliance costs) are obtained from Platts.

Estimating the implicit production subsidies conferred by contingent updating is more complicated. The size of the production subsidy varies with state permit allocation rules, state-specific NOx budgets, annual production levels, and in input-based updating regimes, unit-level heat rates. Individual states allocate their respective NOx "budgets" (listed in Table 3) using formulas of varying complexity.

For each unit, an estimate of the number of future permits earned per unit of current production is constructed using the corresponding state budget E_s, the average ozone season production (or heat input) aggregated across NBP sources in the state, and the specific details of state's permit allocation updating protocols. For example, if NOx permits in state s are allocated based on the average heat input in the preceeding L years, the effect of an incremental increase in current production at firm i in year t on future permit entitlements is assumed to be $h_i \left(\frac{E_s}{H_{st}}\right)$,where h_i measures the fuel inputs required to generate a unit of output at unit i, and H_{st} measures the total quantity of fuel inputs used by NBP sources in state s over the course of the ozone season in year t. This assumes that firms take the size of the subsidy as given.[29] Column 3 of Table 4 summarizes these estimated subsidies, in terms of future permits allocated, per MWh of electricity generated.

In present value dollar terms, the estimated implicit subsidy conferred under this input-based updating regime is:

$$s_{it} = \sum_{l=1}^{L} \frac{\delta_i(l)}{L} \tau_l \left(\frac{h_{it}}{H_t} E_s\right), \tag{12}$$

where τ_l is the expected permit price in l years and $\delta(l)$ is the discount rate applied to benefits accruing l years in the future.[30] In the final column of Table 4, net compliance costs (i.e. explicit compliance costs less the implicit subsidy per MWh) are summarized.[31] These incentives vary sig-

[29] A firm with a dominant market position would want to account for the fact that increasing its fuel consumption would increase H_{st} and thus decrease the size of the subsidy it received for all of its production. Thus we would expect the perceived subsidy would be decreasing in market share.

[30] To construct an estimate these implicit subsidies, the futures price of permits issued l years in the future is used to estimate $\delta_i(l)\tau_l$. In cases where permits did not trade far enough into the future, the market price for the permit vintage farthest in the future was applied to all subsequent vintages..

[31] Rather than assume an arbitrary discount rate, these calculations present undiscounted estimates of the implicit production subsidy. Plant managers presumably discount the value of future permit allocations, so these estimates should be interpreted as generous.

nificantly across facilities. Notably, for several units in allocation updating regimes with relatively low emissions rates, the estimated implicit subsidy exceeds the estimated explicit compliance cost such that the net effect of the NBP on variable operating costs is negative.

Electricity market data

Hourly data summarizing electricity market conditions (including realized and forecast load, real time prices, and day ahead prices) were collected from the New England, New York, and PJM websites. Hourly weather station data were obtained from the National Oceanic and Atmospheric Administration. Additional operating characteristics (including production capacities and fuel characteristics) were obtained from Platts Basecase database.

Fuel price data

Daily spot prices of New York Harbor No. 2 and No. 6 heating oils were obtained from the Energy Information Administration. Daily natural gas spot prices were obtained from Platts.[32] The OTC NYMEX "Look-Alike" contract, the most actively traded coal product, is used as a measure of coal prices.[33]

5 Empirical framework

The unique design of the NBP provides several potentially useful sources of variation. First, the delegation of permit design to state-level agencies has yielded significant interstate variation in permit allocation rules and related incentives. From a research design standpoint, permit allocation design features would ideally have been randomly assigned across electricity producers. Although states' choice of permit allocation regime was not random, interstate variation in permit allocation design is arguably exogenous, in an econometric sense, to firms' short-run production decisions.

[32] For New England I use the natural gas spot prices for Algonquin City Gates and Tennessee, zone 6. For New York gas prices I use Dominion, north point. For PJM I use the Transco Zone 6 non-New York prices.

[33] The NYMEX look-alike price is based on trading using the specifications in the NYMEX futures contract. Originally, NYMEX was predominantly traded "physical" with companies engaging in bilateral agreements. Now a majority of NYMEX trades are transacted OTC then "given up" to the exchange for clearing, which allows market participants to manage credit risk over a larger pool of counterparties.

State level permit allocation design decisions were determined by a variety of factors, including the institutional capacities of the implementing agency and the preferences of politically powerful constituents. Conditional on pre-determined industry and production technology characteristics, the factors that shaped a state's choice of allocation regime should not impact unit-level short run production decisions except through permit allocation incentives.

Second, the seasonal nature of the program's compliance requirements generates useful intertemporal variation. There is considerable overlap in the distribution of hourly load levels, and other observable market conditions across ozone season and off-season (see Table 5). This makes it possible to observe unit-level production decisions in hours that differ in terms of ozone compliance requirements, but share similar market conditions.

Third, a subset of NOx emitting producers supplying the New England electricity market are exempt from the NBP for meteorological reasons. Prevailing wind and weather patterns ensure that emissions from these plants do not contribute significantly to U.S. non-attainment problems . The distributions of operating characteristics that determine short run production decisions in the exempt and NBP regulated sub-populations overlap considerably (see Table 2), making these exempt units a potentially useful control group.

Finally, a majority of states chose to adopt the EPA's recommended permit allocation approach: heat input based updating. Under this regime, the production subsidy varies significantly with fuel efficiency. Input-based updating thus generates interfacility, intra-market variation in production incentives that is independent of variation in explicit compliance costs per unit of production.[34]

5.1 Modeling the data generating process

The model developed in section 2 serves as a good starting point for an empirical model of the process generating observed hourly electricity production decisions. However, a preliminary look at the data suggests that some immediate modifications are needed. First, physical capacity contraints routinely bind in the short-run (units are often observed running at full capacity or

[34]Unit-level fuel efficiency measures are not strongly correlated with NOx emissions rates. The correlation coefficient is 0.69.

shut down completely) so interior solutions are no longer assumed. Second, the vast majority of variation in hourly heat rates occurs across- versus within units (see Appendix X). I therefore assume constant unit-level marginal costs.

Once these two modifications are made, we are left with a model that closely resembles those used to simulate wholesale electricity market outcomes in competitive benchmark analysis (see, for example, Borenstein, Bushnell, and Wolak, 2002 and Wolfram, 1999) and environmental policy simulations (examples include US EPA, 1999 and Burtraw, Palmer and Kahn, 2005). This model predicts that profit maximizing electricity producers follow an on-off strategy, producing at full capacity whenever price exceeds a reservation price set equal to the unit's marginal operating cost. In theory, the introduction of the NBP should increase this reservation price by an amount equal to the unit-specific net compliance cost per MWh.

Are observed production decisions consistent with this prediction? Figure 2 is generated using a small subset of the data collected from a representative unit over a short (three day) period in the ozone off-season.[35] The left panel plots capacity utilization and hourly wholesale electricity prices over these 96 hours. The horizontal line represents the estimated marginal off-season operating cost of \$49/MWh (specific to this unit and time period). The right panel plots capacity factor as a function of the wholesale electricity price less variable operating costs. The thick black line plots the relationship predicted by the model. The thin black line is a local polynomial smooth of the observed data. This figure serves to illustrate how observed hourly production decisions at this representative unit deviate systematically from the predictions of the simple, static model.[36]

Figure 3 conducts a similar exercise using the complete data set. The vertical axis measures capacity factor. The horizontal axis measures price less variable operating costs (not including NBP compliance costs). The solid line in each panel plots the local mean smooth of hourly, unit-level capacity utilization rates on hourly, unit-specific price-cost margins in the ozone off-season.[37]

[35] I chose a period in the ozone off-season in which the wholesale electricity price was vascillating around the unit's theoretical reservation price (i.e. the prevailing fuel price multiplied by the unit's fuel efficiency rating plus variable, non-fuel operating costs).

[36] These supply decisions are observed in the ozone off-season. The average net compliance cost incurred by this unit during ozone season is estimated to be \$3.16. The thick red line illustrates how the introduction of the emissions trading program should, in theory, affect the unit's hourly production decisions.

[37] To generate these figures, I use an Epanechnikov weight function and a rule-of-thumb bandwidth estimator. The smooth is evaluated at 50 points. Price cost margins are calculated by subtracting a unit's fuel costs and

These functions are generated separately for grandfathering and contingent updating regimes, respectively. The broken lines plot the same relationships using data from ozone season (i.e. May to September) when all units are required to purchase permits to offset their NOx emissions.

On average, we should expect that the ozone-season supply functions will lie to the right of their off-season counterparts. Because the average net effect of the NBP on variable operating costs is substantially higher among units operating in grandfathering regimes (see table 4), we should also expect that the NBP-induced shift in the supply curve will be larger in the right panel. The data appear to be consistent with the first prediction, less so with the second.

This preliminary look at the data suggests that the simple, static model poorly approximates the true data generating process. On average, plant managers seem willing to operate in hours when prices fall below marginal operating costs, and are slow to respond when the prices rise above cost. Much of this behavior can likely be explained by production constraints omitted from the model. At the unit level, ramping limits, start up costs, minimum run times, and other intertemporal operating constraints can significantly affect how a plant responds to changing market conditions. At the system-level, transmission constraints, system-security requirements, and other operating protocols can affect which units get called upon to run in a given hour.

The so-called "unit commitment" problem (i.e. the scheduling of electricity production over hours in a day) has been extensively analyzed in the operations research and power systems literature. The problem is difficult to solve because of its large dimension, non-linearity, and large number of constraints (Sheble and Fahd 1990). One formulation of the unit commitment problem, introduced in Appendix 5, helps to convey the complexity of the dynamic optimization faced by plant managers and system operators each day.

Ideally, the estimation framework would accomodate the salient features of the underlying data generating process. Specifying a fully structural econometric model would allow for rich, out of sample, counterfactual welfare analysis. However, this approach would be computationally intensive and would require critical assumptions about the nature of both the system-level and unit-level operating constraints and protocols. Given the limitations of the available data, these

non-fuel variable operating costs per MWh from the hourly real-time wholesale electricity price.

assumptions would be, for the most part, ad hoc and untestable.

In light of these limitations, two alternative empirical strategies are pursued. The first approach is grounded in economic theory only to the extent that theory identifies which variables should affect firms' supply decisions. The second empirical strategy focuses on one specific dimension or margin of the larger unit commitment problem- the decision to begin operating conditional on being shut down- and derives a reduced form that can be implemented empirically as a discrete choice problem.

5.2 A descriptive model of short-run supply decisions

Price-taking producers are assumed to choose output levels based on historic, current, and forecast electricity market conditions, operating costs, and intertemporal operating constraints:

$$cf_{it} = \alpha_i + X'_{it}\gamma_i + NBP_{it} \cdot \tau_t(\pi_1 e_i + \pi_2 s_i) + \varepsilon_{it}, \tag{13}$$

where i indexes the electricity generating unit, t indexes the time period, and cf is the capacity factor (i.e. the percentage of maximum capacity at which the unit is operating). The matrix X includes observable variables that affect short-run supply decisions but are unrelated to the NBP (including fuel prices, day ahead and real time electricity prices, temperature, etc.). The NOx permit price is τ.

The binary indicator NBP_{it} equals one if unit i is obliged to comply with the NBP in period t and zero otherwise. Unit-level emissions rates and implicit subsidies (in terms of future permits earned per unit of electricity generated) are represented by e_i and s_i, respectively. The π coefficients accomodate asymmetric weighting of the explicit compliance costs and implicit production subsidies. The error term ε_{it} is intended to capture the effects of unobserved determinants of the operating decision (including unanticipated deviations from least-cost system dispatch and unit level productivity shocks).[38]

[38] This estimation framework is similar to the approach used by Mansur (2008) to model the short-run production decisions of electricity generating units in the PJM wholesale market.

This descriptive model can be used to assess whether the statistical relationships in the observed data are qualitatively consistent with the theory. However, because the assumed linear conditional mean specification has not been formally derived from the underlying unit commitment problem, any structural interpretation of the coefficient estimates will require additional assumptions. These are introduced in section 6.

5.3 A reduced form model of the participation decision

Consider the unit commitment problem faced by a single electricity generating unit.[39] Let H_i denote the relevant time horizon (measured in hours) for the unit commitment problem solved by unit i. For baseload units that are incapable of responding quickly to changing market conditions, production levels in one hour will constrain production possibilities H hours into the future. Among the most nimble units, $H = 1$ and the production decision reduces to the static benchmark model.

Let q_{it} measure the output level at unit i at the beginning of hour t. The control variable d_{it} measures the change in output at unit i in hour t. The transition equation that determines the evolution of the state variable q_{it} over time is thus $q_{it} + d_{it} = q_{it+1}$. Let D define the decision space. The set of possible production level changes available to unit i in hour t, D_{it}, will depend on time invariant parameters of the operating constraints Γ_i and the state variable q_{it}. The family of unit-specific constraint sets is $\{D_{it}(q_{it}, \Gamma_i)\}$.

Profits earned from the sales of electricity generated at unit i in hour t are:

$$\begin{aligned} \pi_{it} &= (P_{it} - c_i - \theta^t \tau_t e_i + \delta\theta^s \ \tau_t s_i)(q_{it} + d_{it}) - U_i y_{it} - F_i, \ d_{it} \in \{D_{it}(q_{it}, \Gamma_i)\} \\ &\equiv F(q_{it}, d_{it}, X'_{it}) \end{aligned}$$

where the weighting parameters θ are included to allow firms to weigh explicit compliance costs and the implicit subsidy asymmetrically in their production decisions. Let X_{it} denote a matrix of state variables observed by both the plant manager and the econometrician, including fuel

[39] Here I assume that hourly supply decisions are controlled by plant managers insofar as they can submit bids to the independent system operator to achieve their desired production levels.

prices, permit prices, day ahead and real time electricity prices. The superscript $'$ is intended to distinguish this matrix from the larger X matrix in equation [13]. I assume that the plant manager takes X'_{it} as given; this assumption is revisited in the next section.

The binary variable y_{it} equals 1 if the unit turns on in hour t; $y_{it} = 0$. Start-up costs and fixed costs are represented by U and F, respectively. The plant manager's objective is to maximize multi-period profits $\Pi(q_0, X_i^{'}, d) = \sum_{t=0}^{H_i} F(q_{it}, X'_{it}, d_{it})$ subject to the transition function $T(q_t, d_t)$ and the constraint sets $D_{it}(q_{it}, \Gamma_i)$.

Within this framework, theory makes clear predictions about when an inactive unit should start producing. A profit maximizing manager will choose to incur the costs of initiating operations if the profits from doing so exceed the profits associated with remaining out of the market. Focusing exclusively on this participation margin greatly simplifies the derivation of an estimable, more structural model.

This estimation strategy will consider only those unit-hour observations in which $q_{it} = 0$. I use j to index the participation choice situation and t to index the hours relevant to the participation decision: $t = 0...H_i$. I define choice specific value functions $v(y, X''_j)$ to capture the expected profits associated with participation choice:

$$\begin{aligned} v(1, X''_j) &= E_0[F(0, d^*_0(1), X'_j) + \sum_{t=1}^{H} F(q_t, d^*_t(1), X'_j) + \upsilon^1_j] \\ v(0, X''_j) &= E_0[F(0, 0, X'_j,) + \sum_{t=1}^{H} F(q_t, d^*_t(0), X'_j) + \upsilon^0_j. \end{aligned}$$

The i subscripts have been dropped for ease of exposition. The X'' matrix includes all state variables relevant to the participation choice made in hour 0: fuel prices and permit prices which do not vary over the time horizon H, the real time electricity price in hour 0 and the day ahead prices for hours $1..H$. The optimal production choice in hour t conditional on the initial participation choice y is $d^*_t(y)$. A decision specific shock υ^y captures the effects of unobserved factors affecting expected returns to either starting to produce or remaining inactive. These factors could include plant efficiency shocks, unscheduled outtages, operator errors, etc. For simplicity I assume these

shocks are additive, independently distrubuted N(0,σ).

To complete the motivation of the econometric model, I define a latent variable y_{ij}^* to measure the difference in these conditional value functions:

$$y_{ij}^* = v(1, X_{ij}'') - v(0, X_{ij}'') = \alpha_i + \sum_{t=0}^{H} \beta_{ijt}^P P_t - \beta_{ij}^c c_{ij} - \beta^e \tau_j e_i + \ \beta^s \tau_j s_i + \epsilon_{ij}$$

The unit-specific fixed effect α_i captures start up costs, fixed operating costs, and other unobserved, time-invariant factors influencing the participation decision. The electricity price coefficients β_t^P capture the period by period differences in optimal output levels (conditional on the participation decision made in period 0) for a particular unit and choice situation: $\beta_t^P = (q_t(1) + d_t^*(1))$. Let Δ_{ij} measure the effect of the participation decision on production over the time horizon H_i for unit i and choice situation j: $\Delta_{ij} = \sum_{t=1}^{H}(q_{ijh}(1) + d_{ijh}^*(1) - q_{ijh}(0) - d_{ijh}^*(0))$. Note that these Δ_{ij} values vary across units and within units across choice situations; as the decision environment changes (i.e. as fuel prices, permit prices, and electricity prices change), the trajectories of optimal production choices $\{d_t^*(y)\}$ will change. The reduced form parameters β^c, β^e, and β^s coefficients are defined as Δ_{ij}, $\theta_i^t \Delta_{ij}$, and $\theta_i^s \delta_i \Delta_{ij}$, respectively. The random state variable ϵ_{ij} is assumed to be normally distributed (arising from the difference between υ_j^0 and υ_j^1).

The observed binary choice variable y_j serves as an indicator that the latent valye $y_j^* > 0$: $y_{ij} = 1\{y_{ij}^* \geq 1\}$.The probability that an inactive unit i facing choice situation j will begin to operate is given by:

$$\Pr(y_{ij} = 1 | X_{ij}'', \Gamma_i) = \Phi\left(\alpha_i + \sum_{t=0}^{H} \beta_{ijt}^P P_t - \beta_{ij}^c c_{ij} - \beta^e \tau_j e_i + \ \beta^s \tau_j s_i\right).$$

6 Estimation

The data set used in the estimation is hierarchical. Explanatory variables and stochastic components occur at nested micro (i.e. unit-hour) and macro (i.e. electricity generating unit) levels. Each individual unit is observed making production decisions over several thousand hours, so there are sufficient data to carry out a unit-by-unit analysis of relationships between short-run supply

choices and changing electricity market conditions. However, independent variation in permit allocation-related incentives exists only in very limited quantities at the micro-level.[40] The empirical analysis must therefore exploit macro-level variation in unit-level operating characteristics and state-level permit allocation design choices.

Estimation strategies are designed to make efficient use of these hierarchical data without incurring large computational costs. This section describes the two empirical strategies in detail and presents results.

Before turning to the specifics of the estimating equations, two simplifying assumptions deserve mention. First, the contemporaneous permit price τ will be used to proxy for firms' expectations regarding future permit prices. More precisely, I assume that firms value future permit allocations at contemporaneous permit prices and employ a discount rate of zero. This simplifying assumption solves a multicollinearity problem: current and future permit prices are highly correlated. The disadvantage is that the assumption is likely innaccurate. Firms presumably use a non-zero discount rate when valuing benefits accruing in future periods. Also, Table 2 illustrates how futures prices can deviate from spot prices in this permit market. This will have implications for how estimation results should be interpreted.

Second, I treat unit-level emissions rates and heat rates as fixed. No attempt is made to account for the stochastic properties of these unit-specific operating parameters (see Appendix 4). Future work will explore alternative approaches to incorporating this variation.[41]

6.1 Estimating the descriptive model of unit-level production decisions

The γ_i parameters in equation [13] are likely to vary significantly with unobserved factors (such as technical operating constraints). There are sufficient data to estimate these parameters separately

[40] There is some within-unit variation in emissions rates and heat rates (and thus per-MWh compliance costs) across years. In most cases this variation is very small relative to the cross-sectional variation in unit-level emissions rates.

[41] Appendix 4 summarizes how unit-specific emissions rate and heat rate parameters are generated. Joskow and Schmalensee (1985) demonstrate a consistent (adjusted least-squares) technique for using estimated plant operating characteristics as independent variables in crosssection regression analysis.

for each unit, estimating [13] in one step using fully pooled OLS. However, this is cumbersome because it involves thousands of interactions between the variables contained in X_{it} and unit-specific dummy variables. Estimating the model in two steps is computationally easier and has expositional advantages.

In the first step, unit-specific equations are estimated using micro-level data:

$$cf_{it} = \alpha_i + X'_{it}\gamma_i + \pi_{,}\tau_t \cdot D\_OZ_t + \varepsilon_{it}. \tag{14}$$

The dependent variable measures unit-level, hourly capacity factor (on a scale of 0-100). The first two arguments in [14] represent a flexible function that is used to predict firms' operating capacity, given X_{it}, when the cost of emissions is zero. Let $\widetilde{cf}_{it}$ represent this predicted conditional mean value: $\widetilde{cf}_{it} \equiv \alpha_i + X'_{it}\gamma_i$. The X_{it} matrix includes 28 covariates: marginal operating costs, wholesale price (contemporaneous and two lags), forward price (contemporaneous and two leads), daily average price (contemporaneous and lagged), and year fixed effects. With the exception of the fixed effects, these variables enter as third-order polynomials.

The binary variable D_OZ_{it} equals one during ozone season and zero otherwise. Inclusion of $\pi_{,}\tau_t \cdot D\_OZ_t$ in [14] allows predicted production levels to deviate systematically from $\widetilde{cf}_{it}$ during ozone season. These deviations are modeled as unit-specific linear functions of the permit price τ_t .

To address potential endogeneity concerns, ISO-specific hourly demand forecasts are used in some specifications to instrument for electricity prices. Local marginal electricity prices may be endogenous if unobserved supply shocks affect both market clearing prices and unit-level supply decisions. Day ahead demand forecasts should be independent of unobserved supply shocks but highly correlated with realized demand conditions and thus electricity prices in a given hour. Serial correlation is also likely to be a concern given the time series nature of these data. Newey-West autocorrelation consistent standard errors are estimated assuming a twelve hour lag.[42]

In the second stage, the estimated $\pi_{,}$ coefficients are regressed on a constant, a measure of the

[42]Note that this error structure does not account for contemporaneous correlation across the unit-level equations. This is a disadvantage of estimating first stage regression equations separately.

unit-specific NOx emissions rate e_i, and the unit-specific implicit production subsidy s_i. Emissions rates at units exempt from the NOx Budget Program are set to zero. The emissions rate variable is demeaned such that the constant term captures the average relationship between an incremental change in the NOx permit price and operating capacities among units with average NOx emissions rates.

The second stage residuals contain two components: a sampling error component (i.e. the difference between the true value of the estimated dependent variable and the true value) and the idiosyncratic variation that would have obtained regardless of whether the dependent variable was estimated or observed directly. If the sampling variance differs across units, this component will be heteroskedastic. To address this issue, a feasible generalized least squares (FGLS) estimator is used to incorporate information about the variance structure obtained in the first stage estimation (Hanushek,).[43] The weighting matrix used in the second stage is given by:

$$\widehat{\Omega} = \widehat{G} + \widehat{\sigma^2} I, \tag{15}$$

$$\widehat{\sigma}^2 = \frac{\sum_i e_i^2 - \sum_i \epsilon_i^2 + tr(X'X)^{-1}\widehat{G}X}{N-k} \tag{16}$$

where $\widehat{G}$ is the variance covariance matrix from the first stage, e_i^2 are the first stage standard errors of the estimated coefficient that serves as the dependent variable in the second stage, ϵ_i are the residuals from an OLS estimation of the second stage, k is the number of regressors in the second stage and X is the matrix of regressors. This estimator weights more precise first stage estimates more heavily, but only to the degree that sampling error is an important component of the overall second stage residual. Standard errors are also clustered at the facility level to account for the fact that the idiosyncratic component of the error term may be correlated among boilers located at the same facility.

Table 6A summarizes results from estimating [14] using 2SLS. For ease of exposition, this

[43] A more common approach to adjusting standard error estimates when the dependent variable is estimated involves weighting second stage observations using the inverse of the estimated standard error of the dependent variable (Saxonhouse, 1976). However, this assumes that the total residual is heteroskedastic (versus just the component that is explained by sampling error).If the variance due to sampling error accounts for a relatively small fraction of the total residual variance in the second stage, this reweighting can generate misleading standard error estimates.

table reports summary statistics for only a subset of the coefficient estimates: the unit-specific constants, fuel costs ($/MWh), hourly wholesale electricity market price ($/MWh), and the NOx price interacted with the ozone season indicator ($/ton). Corresponding standard errors are also reported.

The marginal cost coefficient is negative on average as expected; the coefficient on (instrumented) electricity price is positive as expected.[44] The coefficient on the NOx permit price interaction is also negative on average. There is cross-sectional variation in the permit price coefficient estimates. The second stage of the estimation seeks to explain some of this variation in terms of observable unit-level characteristics.

Table 7 summarizes results from the second stage of the estimation. The dependent variable is the unit-specific NOx price coefficients obtained in the first stage. The most restrictive specification (1) includes only the demeaned NOx rate and a constant. Results indicate that there is no significant statistical relationship between observed production decisions and ozone season NOx prices among NBP units with average NOx emissions rates. However, the NOX emissions rate coefficient is statistically significant and negative, suggesting that the relationship between NOx permit prices and hourly output is significantly more (less) negative among units with higher (lower) emissions rates. Adding the implicit subsidy variable (specification 2) marginally improves the fit of the model. The coefficient on the implicit subsidy is positive (as expected) but very imprecisely estimated.

Column (3) facilitates a test of whether these coefficient estimates vary systematically with observable unit characteristics.[45] Both the emissions rate and implicit subsidy coefficients are allowed to vary systematically with demeaned unit-level emissions rates and demeaned unit size. In this less restrictive specification, the subsidy coefficient is statistically significant. Of the interaction terms, only the quadratic emissions rate term is significant, suggesting a negative and convex relationship. This specification is also estimated using wholesale electricity prices in the

[44] The wholesale electricity price coefficient is particularly difficult to interpret because this variable is highly correlated with the lagged prices, day ahead lead prices, and higher order terms that are alos included in the regression. Marginal costs also enter as a third order polynomial.

[45] Several alternative specifications were tried but none improved the fit of the model. For example, NOx emissions rates and implicit subsidies were also interacted with heat rates and SO2 emissions rates.

first stage versus instruments. This does not substantively affect the second stage results (reported in column 5).

One potential concern with these results is that the estimated coefficients are capturing spurious correlation. The effect of a change in NOx prices on hourly capacity factors is modeled somewhat crudely within this framework as a vertical shift in $\widetilde{cf}_{it}$ (i.e. the unit-specific function that predicts off-season capacity factor). This shift should only occur in hours when a unit is close to the margin. Otherwise, an incremental change in the permit price should have no effect on short-run production decisions. For many units in the sample, price-cost margins are far below or above zero in majority of hours. In these cases, we should expect to find a very weak or non-existent relationship between permit price variation and unit-level capacity factor (averaged across all hours).

If unit-level price-cost margins are correlated with emissions rates or implicit subsidies, this would confound the interpretation of the estimation results. To address this possibility, two alternative specifications are estimated. First, a (demeaned) measure of the mean squared price-cost margins is added to the model (specification 5). When this variable is controlled for, the constant term becomes statistically significant and negative. The other coefficients are not significantly affected. A second approach involves re-estimating the model using a dataset that omits observations in which units are far from the margin (i.e. price-cost margins exceed \$50/MWh or fall below -\$50). Results are reported in column 6. As expected, all coefficient estimates are larger in absolute value. Both the emissions rate and subsidy coefficient are statistically significant. Taken together, these results suggest that the negative NOx emissions rate coefficient and the positive subsidy coefficient are not capturing the spurious effects of variation in price-cost margins.

To put the results in Table 7 in context, consider a change in the NOx price of \$100/ton. For a unit with an average NOx emissions rate in a grandfathering regime, coefficient estimates reported in column (5) imply an increase in operating costs of approximately \$0.22/MWh.[46] For a unit of average capacity receiving no subsidy, this incremental NOx price change is associated with a decrease in operating capacity of 7% . If this unit is entitled to an implicit subsidy of 2.9

[46] This increase is very small relative to average variable operating costs which exceed \$80/MWh.

lbs/MWh (the average subsidy observed in the data), this decrease falls to 3%. Using the more limited data set that omits hourly observations at units far from the margin, these decreases are 12% and 6%, respectively.

In sum, the statistical patterns found in these data are generally consistent with standard theory. Short-run production decisions appear to respond more negatively to changes in NOx permit prices among units with higher emissions rates and this negative response is more attenuated among units entitled to larger implicit subsidies. This suggests that firms are taking both explicit and implicit permit allocation incentives into consideration when making their hourly production decisions. Although there is some evidence that these relationships vary in predictable ways with NOx emissions rates and plant capacities, these patterns are not robust to changes in specification.

In order to draw more definitive, causal inference from these statistical relationships, additional structure and assumptions are required. The next section summarizes results from estimating the more structured, reduced form specification.

6.2 Estimating the reduced form model of the participation decision

This empirical model conditions on the unit being initially inactive; all unit-hours in which a unit begins an hour with a capacity factor above zero are dropped from the data set used to estimate [5.3]. This amounts to omitting more than 80 percent of observations at coal-fired units, approximately 25 percent of observations at natural gas fired units, and 14 percent of hourly observations at oil-fired units.

Estimation of [5.3] also proceeds in two steps in order to accomodate heterogeneity in the index parameters. Electricity generating units in the study sample vary significantly in terms of the production technologies they use and the wholesale electricity markets they serve. This variation begets unobserved variation in residual variances and the Δ_{ij} parameters. The large quantity of data makes it possible, in principle, to consistently estimate the reduced form coefficients in [**??**]separately for each unit. However, estimating all of these unit-specific index parameters in a single step is difficult given the matrix size limitations of most statistical software. Estimation in two stages is relatively easy to implement.

In the first stage, unit-specific probit equations are estimated:

$$y_{ij} = 1\{\alpha_i + \sum_{t=0}^{H} \beta_{it}^{P} P_t - \beta_i^c c_{ij} - \beta_i^1 \tau_j \cdot D\_OZ_j + \epsilon_{ij}\}.$$

Second stage estimation is very similar to the exercise summarized above. Unit-specific estimates of β_i^1 are regressed on a constant, unit-level emissions rates (interacted with an NBP indicator signaling compliance obligations) and estimated implicit subsidies. An FGLS estimator is used to incorporate information about the variance structure obtained in the first stage.

Interpretation of these estimates is complicated by two identification problems. The first is inherent in all discrete choice models: parameters are identified only up to a scale factor (Maddala 1983). Consequently, comparisons of coefficient estimates across units confound the magnitude of the regression coefficients with residual variation. The second identification problem pertains to the structure of the reduced form estimating equation. Permit price coefficients are confounded with the Δ_{ij} parameters.[47] To identify relationships between permit allocation incentives and short run production decisions, some additional assumptions must be made. These are discussed below.

Table 6B summarizes the estimation results from the first stage. These estimates are generated using a specification that assumes a time horizon of six hours. Hourly demand forecasts are used as instruments for real time and day ahead electricity prices. As expected, estimated marginal cost coefficients are negative on average. Estimated wholesale electricity price coefficients are of the same order of magnitude and are positive on average. The average NOx price coefficient estimate is also negative and highly variable across units.

Estimation results from the second stage are presented in Table 8. The dependent variable is the estimated β_i^1 coefficients from the first stage. The most restrictive specification includes only a constant and demeaned NOx emissions rate. Somewhat surprisingly, the constant term is not statistically significant. This implies that an incremental change in the NOx price has no statistically significant effect on the latent value y^* as defined in [5.3] among units with average

[47] The ratio of the marginal cost and NOx permit price coefficient is identified in principle; scale parameters and Δ parameters cancel out. One approach would involve using this ratio as the dependent variable in the second stage in order to overcome these identification problems. However, this is difficult to implement in practice. Because the distribution of both coefficients include zero, some estimates of this ratio are infinitely large.

NOx emissions rates. However, results indicate that this NOx price coefficient is significantly more (less) negative among units with relatively high (low) emissions rates.

A slightly less restrictive specification (2) allows the NOx price coefficient to vary with the implicit subsidy introduced by contingent allocation updating regimes. The emissions rate coefficient increases slightly in absolute value and remains highly statistically significant. The coefficient on the subsidy variable is positive as expected, but imprecisely estimated and not statistically significant at a 10 percent confidence level.

Alternative specifications were estimated so as to allow the NOx rate and subsidy coefficients to vary with observable unit-level characteristics. The specification that best fit the data is one that allows these coefficients to vary with installed capacity (column 3). Results suggest that the statistically significant and negative coefficient on the NOx emissions rate becomes more (less) negative among units with relatively large (small) operating capacities. Similarly, the statistically significant and positive coefficient on the implicit subsidy becomes more (less) positive among relatively large (small) units.

Columns (4), (5), (6), and (7) report results from estimation exercises intended to test the robustness of these results to varying assumptions about the endogeneity of first stage covariates and the appropriate choice of time horizon. ,Column (4) reports results obtained when electricity prices are assumed to be exogenous in the first stage. Columns (5), (6), and (7) report results under varying assumptions about the time horizon. The NOx rate coefficient is negative and highly statistically significant across all specifications. The statistical significance of the implicit subsidy and the interaction terms are less robust to these specification changes.

One interpretation of these NOx emissions rate and implicit subsidy coefficients is that they provide unbiased measures of direct, causal effects on the relationship between permit prices and short run participation decisions. This interpretation implicitly assumes that the residual variance and the Δ_{ij} parameters are distributed independently of permit allocation incentives. Otherwise, strong correlations between NOx price coefficients estimated in the first stage and emissions rates or implicit subsidies could be spurious.[48] To investigate this possibility, additional specifications

[48]For example, if the Δ_{ij} parameters are systematically smaller among units with higher emissions rates, this would result in a negative NOx emissions rate coefficient in the second stage that has nothing to do with a causal

are estimated. Although the Δ_{ij} parameters and residual variances are not observable, variables that are plausibly strongly correlated with these factors are observed in the data. For example, fast ramp rates should be associated with smaller Δ_{ij} parameters. Larger operating capacities should be associated with larger Δ_{ij} parameters. Residual variances could be expected to vary across regional electricity markets with different dispatch protocols and procedures. Including these variables, and/or interaction terms implicating these variables, did not improve the fit of the model or substantially affect coefficient estimates. These results provide weak support to the aforementioned independence assumptions.

Conditional on these additional identifying assumptions, the estimation results suggest that electricity suppliers do respond to both explicit compliance costs and implicit production subsidies associated with the introduction of this cap-and-trade program. For a variety of reasons (including the discounting of future revenue streams, loss aversion, regulatory risk premiums, etc), we might expect that the implicit subsidy would be weighted less heavily than the explicit and immediate costs of holding permits to offset emissions. This is what we observe in the point estimates across specifications, although these coefficients are quite noisy. Based on these estimation results, we cannot reject the hypothesis that these two operating cost considerations receive equal weight in short run production decisions.

7 Conclusions

Policymakers, industry representatives, and other stakeholders are increasingly interested in understanding how the choice of permit allocation methodology affects permit and product market outcomes in practice. Contingent permit allocation rules simultaneously penalize emissions while rewarding production. A growing number of theoretical papers offer insights into how firms should (in theory) respond to dynamic permit allocation rules. However, the practical implications of contingent allocation updating are not well understood.

A simple partial equilibrium model is used to demonstrate the first order, short run implications of contingent updating vis a vis more traditional permit allocation designs (i.e. auctioning

relationship between high emissions rates and the nature of producers' response to changing NOx permit prices.

and grandfathering). In a first best setting, allocation fails to achieve the efficient, "first-best" outcome. In general, consumer prices are lower, consumption levels are higher, and supply-side emissions abatement costs are greater as compared to grandfathering or auctioning.

These implications, and related results demonstrated elsewhere in both the theory literature and more applied policy simulations, are predicated on the assumption that firms will accurately account for all permit allocation incentives in their short run production decisions. However, there are a variety of reasons why firms might discount, or ignore, the implicit subsidy conferred by contingent allocation updating in practice.

A multi-state emissions trading program offers a unique opportunity to analyze how firms are responding to the incentives created by different emissions permit allocating rules. Two complementary empirical strategies are used to evaluate whether these data are consistent with the theory.

I find that an increase in the emissions permit price has a larger (more negative) effect on the production decisions of units with higher emissions rates in this cap-and-trade program (as theory predicts). This suggests that firms are accounting for the costs of holding permits to offset uncontrolled emissions when making short run supply decisions. The effect of the implicit subsidy introduced by contingent allocation updating is less clear. The data suggest there is considerable heterogeneity in firms' response to this implicit production subsidy. Among average sized producers, there is evidence that production does respond positively to these incentives. Among smaller producers, the evidence is mixed. Finally, I cannot reject the hypothesis that firms place equal weight on explicit environmental compliance costs and implicit production subsidies conferred by the environmental regulation when making short-run supply decisions.

References

(n.d.*a*).

(n.d.*b*).

Bernard, A. L., Fischer, C. & Fox, A. K. (2007), 'Is there a rationale for output-based rebating of environmental levies?', *Resource and Energy Economics* **29**(2), 83–101.

Bohringer, C. & Lange, A. (2005), 'Economic implications of alternative allocation schemes for emission allowances', *The Scandinavian Journal of Economics* **107**(3), 563–581.

Borenstein, S., Bushnell, J. B. & Wolak, F. A. (2002), 'Measuring market inefficiencies in california's restructured wholesale electricity market', *American Economic Review* **92**(5), 1376–1405.

Bunn, D. & Fezzi, C. (2007), 'Interaction of European Carbon Trading and Energy Prices', *SSRN eLibrary* .

Burtraw, D., Palmer, K., Bharvirkar, R. & Paul, A. (2002), 'The effect on asset values of the allocation of carbon dioxide emission allowances', *Electricity Journal* pp. 51–62.

Burtraw, D., Palmer, K. & Kahn, D. (2005), CO2 allowance allocation in the regional greenhouse gas initiative and the effect on electricity investors, Discussion Papers dp-05-55, Resources For the Future.

Cramton, P. & Kerr, S. (2002), 'Tradeable carbon permit auctions: How and why to auction not grandfather', *Energy Policy* **30**(4), 333 – 345.

Dinan, T. & Rogers, D. (2002), 'Distributional effects of carbon allowance trading: How government decisions determine winners and losers', *National Tax Journal* **2**, 199–221.

Environmental Protection Agency, U. (1999), Economic analysis of alternate methods of allocating nox emission allowances, Technical report, ICF Consulting.

Fell, H. (2009), 'Eu-ets and nordic electricity: A cvar analysis', *Energy Journal* .

Fiegenbaum, A. (1990), 'Prospect theory and the risk-return association : An empirical examination in 85 industries', *Journal of Economic Behavior & Organization* **14**(2), 187–203.

Fischer, C. (2003), Output-based allocation of environmental policy revenues and imperfect com-

petition, Discussion Papers dp-02-60, Resources For the Future.

Fischer, C. & Fox, A. K. (2007), 'Output-based allocation of emissions permits for mitigating tax and trade interactions', *Land Economics* **83**(4), 575–599.

Gersbach, H. & Requate, T. (2004), 'Emission taxes and optimal refunding schemes', *Journal of Public Economics* **88**(3-4), 713–725.

Goulder, L. H., Parry, I. W. H., Williams III, R. C. & Burtraw, D. (1999), 'The cost-effectiveness of alternative instruments for environmental protection in a second-best setting', *Journal of Public Economics* **72**(3), 329–360.

Jensen, J. & Rasmussen, T. N. (2000), 'Allocation of co2 emissions permits: A general equilibrium analysis of policy instruments', *Journal of Environmental Economics and Management* **40**(2), 111–136.

Montgomery, D. W. (1972), 'Markets in licenses and efficient pollution control programs', *Journal of Economic Theory* (3), 395–418.

National Commission on Energy Policy (2007), Allocating allowances in a greenhouse gas trading system, Staff paper, NCEP.

Neuhoff, K., Martinez, K. K. & Sato, M. (2006), 'Allocation, incentives and distortions: the impact of eu ets emissions allowance allocations to the electricity sector', *Climate Policy* **6**(1), 73–91.

Quirion, P. & Demailly, D. (2006), 'CO2 abatement, competitiveness and leakage in the European cement industry under the EU ETS: grandfathering versus output-based allocation', *Climate Policy* **6**, 93–113.

Quirion, P. & Demailly, D. (2008), Changing the allocation rules in the eu ets: Impact on competitiveness and economic efficiency, Working Papers 2008.89, Fondazione Eni Enrico Mattei.

Sijm, J., Neuhoff, K. & Chen, Y. (2006), Co2 cost pass through and windfall profits in the power sector, Cambridge Working Papers in Economics 0639, Faculty of Economics, University of Cambridge.

Sterner, T. & Muller, A. (2008), 'Output and abatement effects of allocation readjustment in permit trade', *Climatic Change* **86**(1-2), 33–49.

U.S. Environmental Protection Agency (2009), The United States Environmental Protection Agency's Preliminary Analysis of the Waxman-Markey Discussion Draft in the 111th Congress, The American Clean Energy and Security Act of 2009, Technical report, http://www.epa.gov/climatechange/economics/economicanalyses.html.

Wolfram, C. D. (1999), 'Measuring duopoly power in the british electricity spot market', *American Economic Review* **89**(4), 805–826.

Zachmann, G. & von Hirschhausen, C. (2008), 'First evidence of asymmetric cost pass-through of eu emissions allowances: Examining wholesale electricity prices in germany', *Economics Letters* **99**(3), 465 – 469.

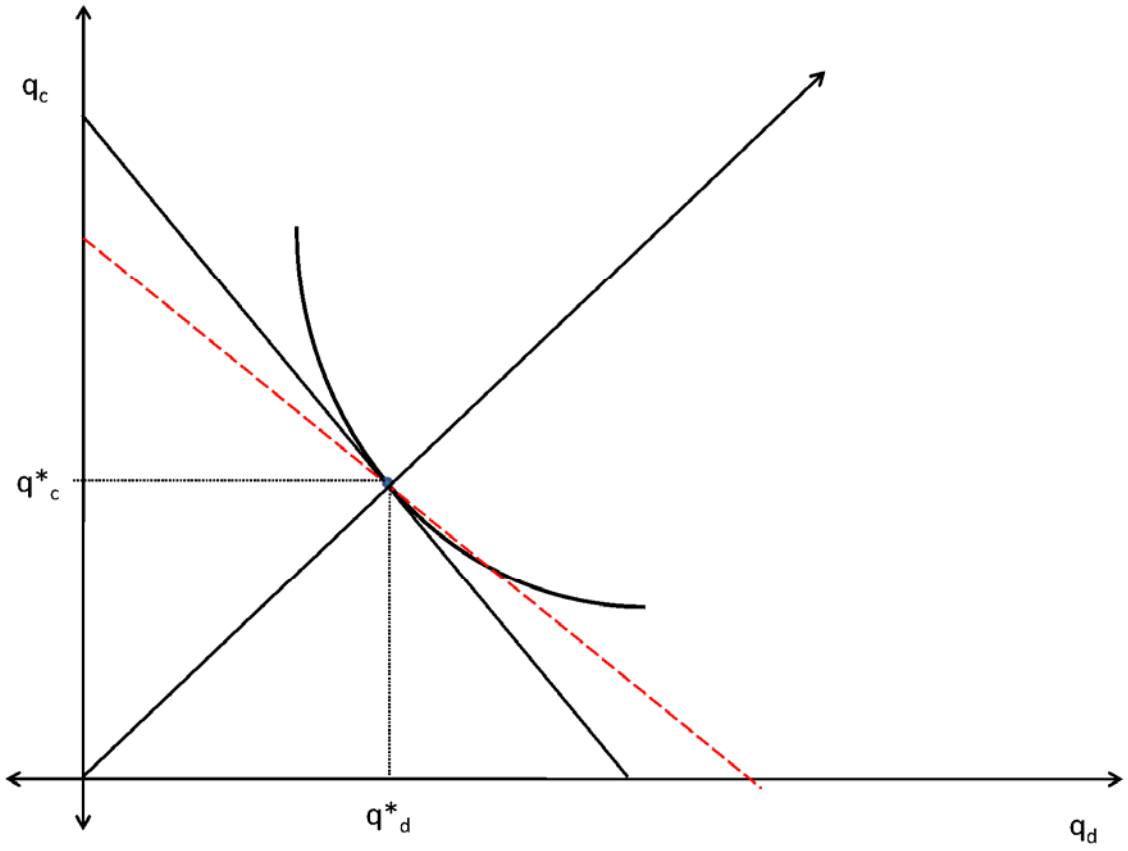

Figure 1 : Emissions constrained social welfare maximizing outcome

Notes : The downward sloping solid line represents the emissions constraint. The socially optimal allocation of production occurs at the point where the emissions constraint is just tangent to a level set of the economic surplus function. This point is intersected by the vector from the origin. The broken line, with a slope of -1, connects all points that correspond to an aggregate output quantity equal to that associated with the optimum outcome. Points lying on the emissions constraint above (below) the optimal point are associated with more (less) consumption and more (less) supply side abatement than is consistent with welfare maximization.

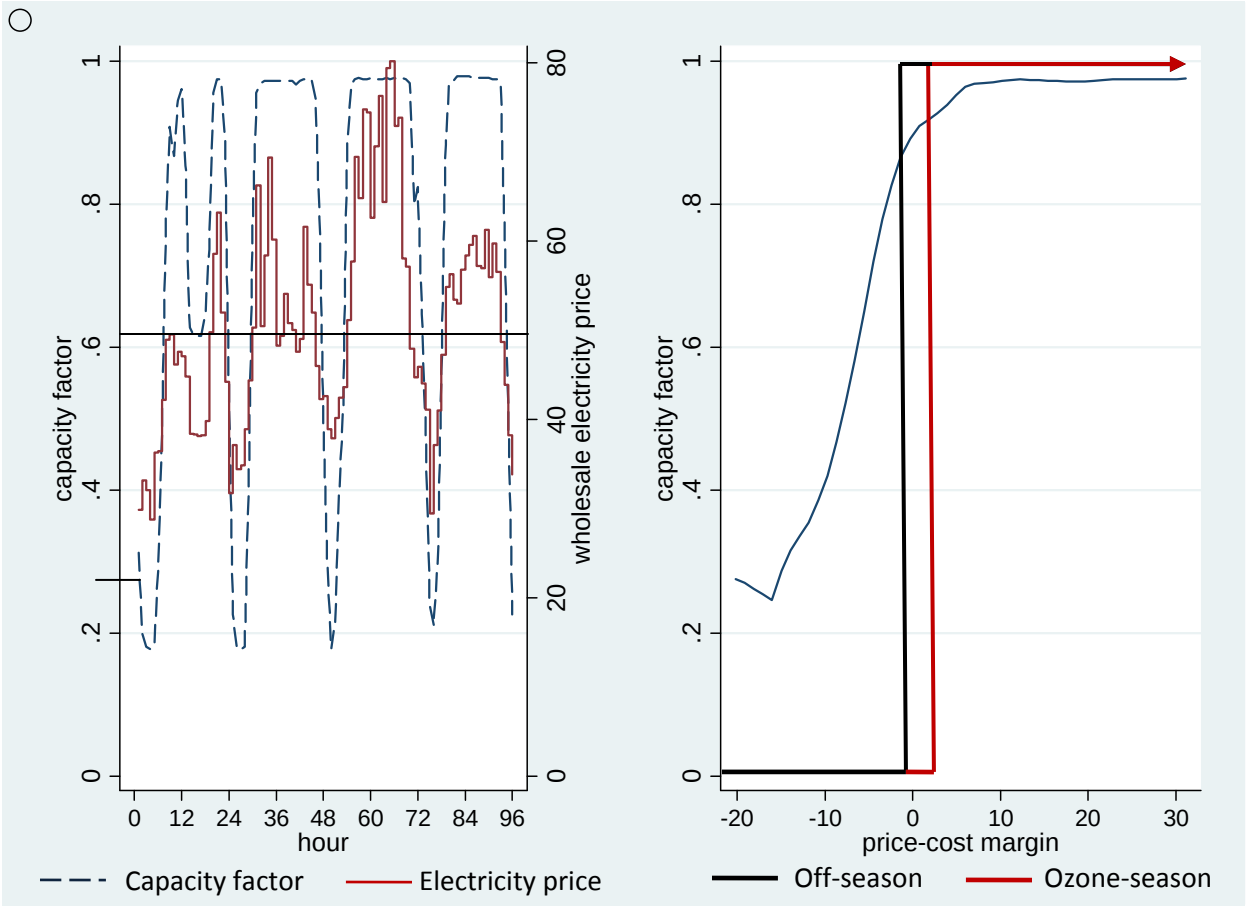

Figure 2 : Hourly Production Decisions at a Representative Unit

Notes : Hourly production decisions at a single unit (measured as capacity factor) and the corresponding hourly wholesale electricity price over a four day period are plotted in the left panel. The horizontal line represents the theoretical reservation price during these off-season hours (i.e. the unit's constant marginal operating cost). This cost of $49/MWh is estimated using the fuel input prices that prevailed in this four day period, the unit-specific heat rate, and other variable (non-fuel) operations costs. The thin black line in the right panel plots these same data in capacity factor, price-cost margin (i.e. the hourly wholesale electricity price less the marginal operating costs incurred) space. This is a mean smooth of capacity factor on price-cost margins. The thick black line represents the on-off production protocol implied by the benchmark model of a profit maximizing, price taking producer. Comparing these two functions helps to illustrate how observed production decisions deviate systematically from the predictions of the simple, static model. Finally, the thick red line represent the expected impact of the NBP on these short-run supply functions. In theory, the unit's reservation price should increase by an amount equal to the net compliance costs per MWh.

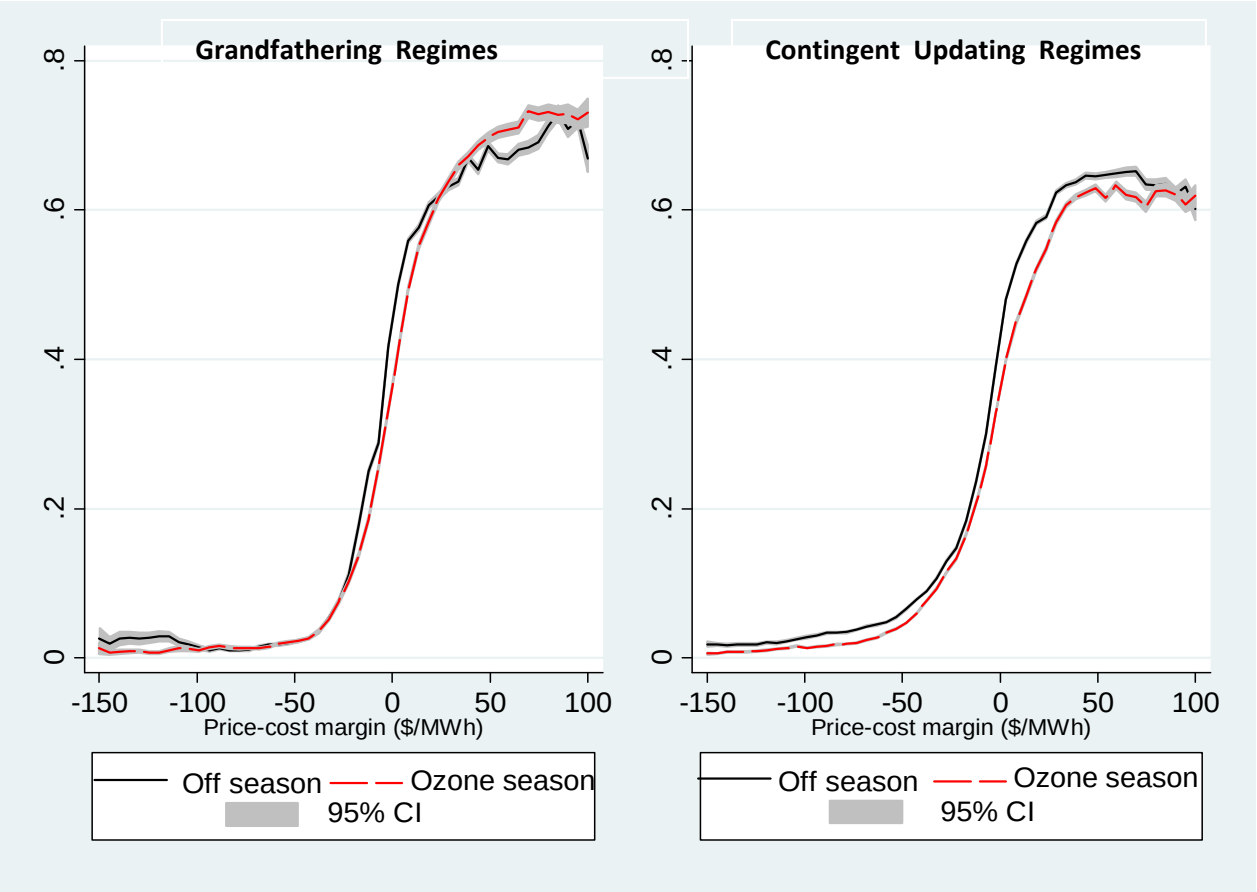

Figure 3 : Electricity Supply Decisions in Grandfathering and Contingent Allocation Updating Regimes

Notes : These figures plot a local mean smooth of hourly, unit-level capacity factors on hourly, unit-specific price-cost margins, where costs include fuel costs and other variable operating costs but exclude any costs or implicit subsidies associated with the NOx Budget Program. The shaded regions represent 95 percent confidence intervals. These graphs are generated using hourly from almost 600 electricity generating units over the study time period (2003-2006), excluding the top and bottom five percentiles of observations. I use an Epanechnikov weight function and rule-of-thumb bandwidth estimator. The right panel summarizes the production decisions at 73 units operating in states where NOx permits are grandfathered. The left panel summarizes production decisions at 515 units operating in states where permit allocations are periodically updated based on lagged input or output choices.

Table 1 : Permit allocation regime chosen by New York, New England, and Mid-Atlantic states

Electricity market	State	Annual state NOx budget for Electricity generating units (tons NOx)	Chosen permit allocation regime
New England	CT	4,253	Output-based updating
	MA	12,861	Output-based updating
	ME	N/A	N/A
	NH	N/A	N/A
	RI	936	Grandfathering
	VT	N/A	N/A
New York	NY	30,405	Input-based updating
PJM	DC	233	Grandfathering
	DE	4,463	Grandfathering
	MD	14,520	Grandfathering
	NJ	8,200	Output-based updating
	PA	47,244	Input-based updating
	VA	17,091	Input-based updating

Notes: Annual, state-level NOx budgets do not change over the study period. Among states that chose contingent updating, implementation details vary considerably. For example, whereas some states update annually, others update in three or four year blocks.

Table 2 : Operating summary statistics by permit allocation regime : CENSUS

Allocation regime	# Units	Summer Capacity (MW)	Off-season capacity factor	Heat rate* (btu/kWh)	Ozone season NOx rate* (lbs NOx/MWh)
Input-based updating	362	155 (206)	19% (26%)	11,384 (2,444)	2.01 (2.04)
Output-based updating	153	134 (176)	10% (18%)	12,857 (3,090)	2.82 (4.54)
Grandfathering	70	200 (182)	21% (26%)	12,080 (3,447)	3.07 (3.60)
Exempt	25	181 (234)	23% (29%)	11,592 (5,388)	1.57 (4.67)

Notes: Standard deviations in parentheses. Summary statistics are generated using data from 610 fossil fuel-fired electricity generating units supplying the New York, New England, or PJM markets during the study period (2003-2006). Self generating and co-generating units are excluded from the sample.
* Emissions rate and heat rate summary statistics are weighted by installed capacity.

Table 3 : NOx Allowance Prices 2003-2006 (Nominal $/ton)

Permit vintage	Transaction year			
	2003	**2004**	**2005***	**2006***
2003	**$3682**	$1906	-	-
2004	$3163	**$2250**	$2180	$1507
2005	$2204	$3432	**$2771**	$1507
2006		$2951	$3018	**$1842**
2007		$2665	$2705	$1750
2008		$2705	$2299	$1570
2009		$2314	$2232	$1518

Notes: This table reports average annual permit prices by NOx permit vintage. Contemporaneous permit prices appear in bold. The asterisk denotes years in which the progressive flow control (PFC) constraint was binding. The PFC ratio was 0.25 and 0.27 in 2005 and 2006, respectively; banked permits are traded at a discount in these years. Permit price data were purchased from Evolution Markets LLC.

Table 4 : Estimated NBP compliance costs and production incentives by allocation regime

Allocation regime	(1) NOx permit costs per MWh generated	(2) (1) as a percentage of off-season variable operating costs	(3) Future permits allocated (tons) per MWh generated	(4) Estimated net compliance cost per MWh*
Input-based updating	$4.89 ($4.69)	6.4% (6.0%)	0.001 (0.001)	$1.40 ($4.69)
Output-based updating	$7.46 ($8.24)	6.5% (5.9%)	0.002 (0.001)	$1.45 ($8.24)
Grandfathering	$6.49 ($9.03)	8.1% (6.8%)	0	$6.49 ($9.03)
Exempt	$0	--	0	$0

Notes: Standard deviations in parentheses. Summary statistics are generated using data from 610 fossil fuel-fired electricity generating units supplying the New York, New England, or PJM markets during the study period (2003-2006). Self generating and co-generating units are excluded from the sample. Averages across unit-years are reported; standard deviations are in parentheses.

* To calculate net compliance costs, future permits allocated per unit of output are valued using futures permit prices. This value is then subtracted from the explicit compliance cost per MWh (i.e. the product of the unit-specific emissions rate and the spot NOx permit price).

Table 5 : Regional wholesale electricity prices and electricity demand by season

Regional electricity market	Hourly electricity demand (MW)		Wholesale electricity price ($/MWh)	
	Off-season	Ozone season	Off-season	Ozone season
New England (NEPOOL)	14,603 (2,513)	15,344 (3,416)	$60.88 ($29.34)	$57.98 ($34.88)
New York (NYISO)	17,547 (2,768)	19,151 (3,936)	$68.00 ($96.97)	$70.74 ($78.17)
Mid-Atlantic (PJM)	29,214 (8,787)	31,694 (9,957)	$52.03 ($33.84)	$55.41 ($43.34)

Notes: Standard deviations in parentheses. This table summarizes observed prices and load levels over the 35,040 hours in the data. Both prices and load levels are similarly distributed across seasons.

Table 6A : Summary of first stage estimates: Dependent variable is hourly capacity factor

Covariate	Average point estimate	Standard deviation	Average standard error
Unit-level constant	25.24	93.49	62.66
Marginal operating cost ($/MWh)	-9.24	12.93	4.18
Local wholesale marginal price ($/MWh)	3.04	17.76	9.39
NOx price * Ozone indicator ($/ton)	-0.003	0.25	0.0008
Average # observations/unit		15,087	

Notes: The unit of analysis is a unit-hour. The dependent variable is the capacity factor (measured on a scale of 1-100). For ease of exposition, only the higher order cost and price variables, lagged prices, futures prices, and year dummies are omitted from this table.

Table 6B : Summary statistics for the first stage estimates

Covariate	Average point estimate	Standard deviation	Average standard error
Unit-level constant	-2.98	1.78	1.69
Marginal operating cost ($/MWh)	-0.14	0.03	0.02
Local wholesale marginal price ($/MWh)	0.06	0.09	0.05
NOx price * Ozone indicator ($/ton)	-0.0004	0.018	0.0001
Average # observations/unit		13,945	

Notes: The unit of analysis is a unit-hour. The dependent variable is the binary participation indicator. Each unit-level regression includes: a unit-level fixed effect, marginal operating costs, contemporaneous wholesale price , hourly forward prices for 1 through 6 hours forward, the NOx permit price interacted with the ozone season indicator and year fixed effects. Real time and day ahead electricity prices are instrumented for using hourly demand forecasts.

Table 7 : Descriptive model : Second stage estimation results

Specification	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	-0.00	-0.02	-0.05	-0.03*	-0.07**	-0.12**
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
NOx rate	-14.56**	-16.27***	-30.02***	-28.46***	-31.35***	-49.85***
	(5.65)	(5.59)	(10.90)	(7.24)	(11.75)	(15.93)
Estimated subsidy		10.82	23.42**	17.06**	26.00**	41.31**
		(9.84)	(11.00)	(7.96)	11.46	(17.83)
$(\text{NOx rate})^2$			1792.17**	1413.27***	2033.56	2143.83
			825.67	467.35	887.56	1316.53
NOx rate * Estimated subsidy			1279.88	550.55	635.06	1488.16
			(1831.50)	(1359.16)	(1727.62)	(7748.64)
NOx rate * capacity			-0.05	-0.09*	-0.08	-0.22**
			(0.05)	(0.05)	(0.06)	(0.09)
Estimated subsidy * capacity			0.08	0.02	0.06	-0.09
			(0.05)	(0.05)	(0.06)	(0.09)
Mean price-cost margin					-0.0001	
					(0.0003)	
R^2	0.09	0.10	0.08	0.10	0.09	0.13
N	553	553	553	553	553	553
First stage IV	Y	Y	Y	N	Y	Y

Notes: The unit of analysis is an electricity generating unit. The dependent variable is the estimated NOx price effect (in ozone season) from the unit-level linear regression estimation. In the first stage, forecast demand is used to instrument for local marginal electricity prices unless otherwise indicated. Emissions rates (demeaned) and estimated subsidies are measured in tons of NOx per MWH. Production capacity (also demeaned) is measured in MW. FGLS standard errors (in parentheses) are clustered at the facility level. To estimate specification (6), all unit-hours in which the estimated price margin is greater than $50 or less than -$50 are dropped in the first stage.

* Statistically significant at the 10 percent level.

** Statistically significant at the 5 percent level.

*** Statistically significant at the 1 percent level.

Table 8 : How do NOx permit prices affect the participation decision? Second stage estimation results

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.000 (0.000)	-0.003 (0.002)	-0.004*** (0.001)	-0.002 (0.001)	-0.002 (0.002)	-0.003** (0.001)	-0.002 (0.002)
NOx rate (tons/MWh)	-2.21*** (0.80)	-2.44*** (0.82)	-3.04*** (0.44)	-2.05*** (0.49)	-2.87*** (0.54)	-3.21*** (0.44)	-2.97*** (0.65)
Estimated subsidy (tons/MWh)		1.69 (1.23)	1.92*** (0.70)	1.45* (0.76)	1.48* (0.84)	1.65** (0.70)	1.47 (0.99)
NOx rate * capacity			-0.008** (0.002)	-0.002 (0.003)	-0.003 (0.004)	-0.008** (0.003)	-0.014*** (0.01)
Estimated subsidy * capacity			0.001* (0.004)	0.007* (0.004)	0.004 (0.004)	-0.001 (0.004)	0.010*** (0.005)
R^2	0.09	0.10	0.11	0.08	0.10	0.13	0.06
N	550	550	550	553	550	553	553
Assumed time horizon (hours)	6	6	6	6	2	4	8
First stage IV?	Y	Y	Y	N	Y	Y	Y

Notes: The unit of analysis is an electricity generating unit. The dependent variable is the coefficient on the interaction between the NOx permit price and ozone season indicator from the first stage of the estimation. First stage probit equations also include a constant, unit-specific marginal operating cost, year indicators, and instruments for current and future wholesale electricity prices (unless otherwise indicated). These first stage equations are estimated under different assumptions about the relevant time horizon (measured in hours). Emissions rates (demeaned) and estimated subsidies are measured in tons of NOx per MWH. Production capacity (also demeaned) is measured in MW. FGLS standard errors (in parentheses) are also clustered at the facility level.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

Appendix 1

Using the system of first order conditions, solve for equilibrium quantities and permit price in terms of the model parameters:

$$Q^* = \frac{Ee_c c_d - 2ae_c e_d + Ee_d c_c + ae_c^2 + ae_d^2}{be_c^2 - 2be_c e_d + be_d^2 + e_c^2 c_d + e_d^2 c_c} \tag{1}$$

$$q_c^* = \frac{bEe_c - bEe_d - ae_c e_d + Ee_c c_d + ae_d^2}{be_c^2 - 2be_c e_d + be_d^2 + e_c^2 c_d + e_d^2 c_c} \tag{2}$$

$$q_d^* = \frac{bEe_d - bEe_c - ae_c e_d + Ee_d c_c + ae_c^2}{be_c^2 - 2be_c e_d + be_d^2 + e_c^2 c_d + e_d^2 c_c} \tag{3}$$

$$\tau^* = \frac{a(e_c c_d + e_d c_c) - E(bc_c + bc_d + c_c c_d)}{b(e_d - e_c)^2 + e_c^2 c_d + e_d^2 c_c} \tag{4}$$

Substituting these expressions into the economic surplus function and differentiating with respect to the emissions constraint:

$$\frac{\partial S^*}{\partial E} = \frac{a(e_c c_d + e_d c_c) - bE(c_d + c_c + c_c c_d)}{b(e_c - e_d)^2 + e_c^2 c_d + e_d^2 c_c} = \tau^*. \tag{5}$$

By equation [**??**] , τ^* is also the marginal cost of reducing emissions on the supply side. Thus, in the first best case, marginal abatement costs are set equal across all producers and across the supply and demand side of the product market.

Appendix 2 :

Profit maximization under symmetric updating yields lower consumer prices and higher supply side abatement costs as compared to the first best case. To see this, I solve for the equilibrium quantities and permit price in terms of the model parameters:

$$Q^{S\_UP} = \frac{Ee_c c_d + Ee_d c_c - 2ae_c e_d - sEc_d - sEc_c + ae_c^2 + ae_d^2}{be_c^2 - se_c c_d - se_d c_c - 2be_c e_d + be_d^2 + e_c^2 c_d + e_d^2 c_c} \tag{6}$$

$$q_c^{S\_UP} = \frac{bEe_c - bEe_d - ae_c e_d - se_c e_d + Ee_c c_d + ae_d^2 + se_d^2}{be_c^2 - 2be_c e_d + be_d^2 + e_c^2 c_d + e_d^2 c_c} \tag{7}$$

$$q_d^{S\_UP} = \frac{bEe_d - bEe_c - ae_c e_d - se_c e_d + Ee_d c_c + ae_c^2 + se_c^2}{be_c^2 - 2be_c e_d + be_d^2 + e_c^2 c_d + e_d^2 c_c} \tag{8}$$

$$\tau^{S\_UP} = \frac{a(e_c c_d + e_d c_c) - E(bc_c + bc_d + c_c c_d) + s(e_c c_d + e_d c_c)}{b(e_2 - e_1)^2 + e_1^2 c_2 + e_2^2 c_1} > \tau^* \tag{9}$$

Assuming the implicity production subsidy is strictly greater than zero, the permit price under updating (and thus the marginal abatement cost on the supply side) is higher under updating versus the benchmark case. It can also be shown that consumption levels will be higher under symmetric updating:

$$
\begin{aligned}
Q^{S\_UP} - Q^* &= \frac{Ee_cc_d + Ee_dc_c - 2ae_ce_d + ae_c^2 + ae_d^2 - sEc_d - sEc_c}{be_c^2 - 2be_ce_d + be_c^2 + e_c^2c_d + e_d^2c_c - se_cc_d - se_dc_c} - && (10)\\
&\quad \frac{Ee_cc_d - 2ae_ce_d + Ee_dc_c + ae_c^2 + ae_d^2}{be_c^2 - 2be_ce_d + be_d^2 + e_c^2c_d + e_d^2c_c} && (11)\\
&= \left(Ee_cc_d - sEc_d - 2ae_ce_d - sEc_c + Ee_dc_c + ae_c^2 + ae_d^2\right) * && (12)\\
&\quad \left(Ee_cc_d - 2ae_ce_d + Ee_dc_c + ae_c^2 + ae_d^2\right) && (13)\\
&= \left(E(e_cc_d + e_dc_c) + a(e_c - e_d)^2 - E(sc_c + sc_d)\right)\left(E(e_cc_d + Ee_dc_c) + a(e_c - e_d)^2\right) && (14)
\end{aligned}
$$

Note that the average emissions rate per unit of output when emissions are unconstrained is $\frac{e_cc_d+e_d+c_c}{c_c+c_d}$. Assuming the cap is binding, $s < \frac{e_cc_d+e_d+c_c}{c_c+c_d}$. This implies that $e_cc_d + e_dc_c > sc_c + sc_d$. Thus, $Q^{OBA} > Q^*$.

Appendix 3

Profit maximization under symmetric updating yields lower consumer prices and higher supply side abatement costs as compared to the first best case. To see this, I solve for the equilibrium quantities and permit price in terms of the model parameters:

$$
\begin{aligned}
Q^{A\_UP} &= \frac{Ee_cc_d + Ee_dc_c - 2ae_ce_d + ae_c^2 + ae_d^2 + e_c^2s_d + e_d^2s_c - e_ce_ds_c - e_ce_ds_d}{be_c^2 + be_d^2 + e_c^2c_d + e_d^2c_c - 2be_ce_d} && (15)\\
q_c^{A\_UP} &= \frac{bEe_c - bEe_d - ae_ce_d + Ee_cc_d + ae_d^2 + e_d^2s_c - e_ce_ds_d}{be_c^2 - 2be_ce_d + be_d^2 + e_c^2c_d + e_d^2c_c} && (16)\\
q_d^{A\_UP} &= \frac{bEe_d - bEe_c - ae_ce_d + Ee_dc_c + ae_c^2 + e_c^2s_d - e_ce_ds_c}{be_c^2 - 2be_ce_d + be_d^2 + e_c^2c_d + e_d^2c_c} && (17)\\
\tau^{A\_UP} &= \frac{a(e_cc_d + e_dc_c) - E(bc_d + bc_c + c_cc_d) + be_cs_c}{b(e_d - e_c)^2 + e_c^2c_d + e_d^2c_c} - && (18)\\
&\quad \frac{be_cs_d - be_ds_c + be_ds_d + e_cc_ds_c + e_dc_cs_d}{b(e_d - e_c)^2 + e_c^2c_d + e_d^2c_c} && (19)
\end{aligned}
$$

If $\frac{e_c}{e_d} = \frac{s_d}{s_c}$, note that $e_d^2s_c = e_ce_ds_d$ and $e_c^2s_d = e_ce_ds_c$. Thus, if the production subsidies are set such that the subsidy per unit of emissions are equal, the first-best production quantities and product price is achieved. Although the product price in this case will be equal to P^*,note that the permit price will exceed τ^* (to compensate for the relatively high subsidy paid to the more polluting firm). If the subsidy per unit of pollution is higher for cleaner firms (as is the case when production subsidies are symmetric), $e_d^2s_c > e_ce_ds_d,\ e_c^2s_d > e_ce_ds_c$, and consumption exceeds first best levels. Alternatively, if the subsidy per unit of pollution is higher for the more polluting firm, the reverse will be true.

Appendix 4 : Unit-level operating characteristics

A more complete appendix 4 will be included in future versions of the paper:

Emissions rates and heat rates are observed at the unit-level over thousands of hours. Table A1 decomposes the observed variation in hourly emissions rates and heat rates. Although the vast majority of the observed variation is between units, there is some within-unit variation in both operating attributes. This within-unit variation can be a result of production choices made by the plant manager (i.e. choice of fuel characteristics, the capacity at which the plant is running, combustion tuning, etc.) or it can be affected by purely exogenous factors (such as ambient temperature).
One might be concerned if operational changes made to reduce a unit's NOx emissions rate have significant impacts on heat rates. In fact, there is little evidence to suggest that emissions rates and heat rates are systematically correlated once other factors- such as capacity factor and temperature- are controlled for.

Unit-specific, season-specific (i.e. ozone season or off-season) measures of these two important operating parameters are obtained by estimating regression equations separately for each unit and season. The following general specification is estimated:

$$z_{its} = \alpha_{is} + X'_{its}\beta_{is} + \varepsilon_{its},$$

where z_i is the operating characteristic of interest (i.e. heat rate or emissions rate) and X is a matrix of variables that affect a unit's operational performance (ambient temperature and capacity factor). Because the variables in X are demeaned, the intercept a_i can be interpreted as the unit-specific. season-specific value of z under average operating conditions for unit i. Table A2 summarizes these results.

These equations are estimated using all available data and data from hours when units' capacity factors exceeded 5%. As electricity generating units increase production, they run more efficiently and reduce fewer emissions per unit of output.

Appendix 5 : The Unit Commitment Problem

The unit commitment problem (UCP) confronting electricity system operators has been extensively studied in the operations research literature. The problem obtains decisions for a sequence of time periods (typically 24 hours of a day). What happens in one hour affects what can happen in subsequent hours; optimal unit commitment in one hour is not independent of solutions in other hours. The following UCP specification was adapted from Hobbs et al. (2001)

$$
\begin{aligned}
& \min \ \sum_{it} z_{it} F_i + \sum_{it} q_{it} c_i + \sum_{it} y_{it} U_i \\
& s.t. \\
\sum_i q_{it} &= D_t \quad \forall t \\
\sum_i r_{it} &= SD_t \quad \forall t \\
q_{it} &\geq z_{it} MIN_i \quad \forall i,t \\
q_{it} + r_{it} &\leq z_{it} MAX_i \quad \forall i,t \\
r_{it} &\leq z_{it} MAXSP_i \quad \forall i,t \\
q_{it} &\leq q_{it-1} + MxINC_i \quad \forall i,t \\
q_{it} &\geq q_{it-1} + MxDEC_i \quad \forall i,t \\
z_{it} &\leq z_{it-1} + y_{it} \quad \forall i,t \\
z_{it} &\geq z_{it-1} = x_{it} \quad \forall i,t \\
& \sum a_{ki}(q_{it}) \leq MxFlow_{it} \ \forall \ k,i,t
\end{aligned}
$$

where the decision variables are:

q_{it}	the MW produced by unit i in hour t
r_{it}	the MW of spinning reserves provided by unit i in hour t
z_{it}	A binary variable indicating whether unit i is dispatched in hour t.
y_{it}	A binary variable indicating whether unit i starts up (having been switched off in hour $t-1$) in hour
x_{it}	A binary variable indicating whether unit i shuts down (having been on in hour $t-1$) in hour t.

The parameters of the problem are:

D_t	Electricity demand in period t.
SD_t	Demand for spinning reserves in period t.
F_{it}	Fixed costs of operating unit i in period t (measured in \$/hour).
c_{it}	Variable operating costs at unit i in period t (measured in \$/MWh).
U_{it}	The costs of starting up unit i (measured in \$).
MAX_i	The maximum production capacity of unit i.
MIN_i	The minimum production capacity of unit i (assumed to be 0 for all i)
$MxINC_i$	Maximum ramp-up rate for increasing production at unit i
$MxDEC_i$	Maximum ramp-down rate for decreasing production at unit i.
a_{ki}	Linearized coefficient relating bus i injection to line k flow.
$MxFlow_{it}$	Maximum MW flow on line k in period t.

In the restructured electricity markets considered in this paper, market participants submit bids to an independent system operator (ISO). Unit-level production activities are coordinated via a two-settlement market system. A day ahead forward market schedules resources and determines hourly prices for the following day; a balancing market ensures that supply meets fluctuating demand in real time. Once generators have submitted their supply bids, independent system operators identify unit commitment schedules to minimize the cost of meeting electricity demand subject to thousands of unit-level and system-level operating constraints.[1] If firms act as price-takers, submitting bids that reflect true costs and operating constraints is a likely to be a profit maximizing

[1] The different regional electricity markets require generators to submit bids in different forms. For example, the New York ISO separates energy bids into start-up cost curves and an energy curve which can be a three part step function or a 6 segment piecewise linear curve.

strategy (G.Gross and D.J.Finlay, "Optimal bidding strategies in competitive electricity markets," Proceedings of lp Power Systems Compuwion Conference (PSCC'96), Dresden, August 19-23, 1996, pp.815-823). Bidding strategies become more complicated in imperfectly competitive market settings.