

Emission trading for air pollution hot spots: getting the permit market right

Werner Antweiler¹

Received: 28 August 2015 / Accepted: 1 December 2015 / Published online: 29 December 2015
© Society for Environmental Economics and Policy Studies and Springer Japan 2015

Abstract Conventional emission permit markets are inefficient for non-uniformly mixed pollutants that create geographic ‘hot spots’ of different ambient emission concentrations and environmental damage. Economically efficient ambient concentration contribution markets involve difficult interactions among multiple markets that makes them practically infeasible. Extending economic theory by Muller and Mendelsohn (Am Econ Rev 99(5):1714–1739, 2009. doi:10.1257/aer.99.5.1714) and others about ‘getting the prices right’ through bilateral trading ratios, this paper introduces theoretical simplifications and a novel type of single permit market with a hybrid price-quantity instrument that addresses the dual heterogeneity of firm-specific abatement costs and regional variation in damage. This paper shows how to ‘get the market right’ robustly through simplicity, liquidity, and gradualism. Analytic solutions and simulation results demonstrate the feasibility of the novel market concept. Also discussed is the potential applicability of the market design to interstate trading in the United States in the wake of the recently implemented *Cross-State Air Pollution Rule*.

Keywords Emission permit markets · Air pollution · Non-uniformly mixed pollutants · Environmental policy

This paper has benefited greatly from presentations at the University of British Columbia, the University of Kiel, the University of Cologne, York University, the University of Maryland, as well as presentations at the Canadian Economics Association conference, and the Canadian Resource and Environmental Economics study group meeting in 2012. I am grateful to all the audiences and individuals who have provided feedback and comments.

✉ Werner Antweiler
werner.antweiler@ubc.ca

¹ Sauder School of Business, University of British Columbia, 2053 Main Mall,
Vancouver BC V6T 1Z2, Canada

1 Introduction

While emission permit trading systems have grown in use over the last decades, their application potential remains somewhat limited because such trading systems work well only for uniformly mixed pollutants such as greenhouse gases or widely dispersed pollutants. Conventional emission permits for non-uniformly mixed pollutants are economically inefficient because emission concentrations vary strongly across receptor locations. ‘Hot spots’ are localized areas with high concentrations of a pollutant. The efficacy of any policy intervention depends crucially on how it targets these ‘hot spots.’ Interventions have to target high-damage ‘hot spots’ more aggressively than low-damage regions.

Montgomery (1972) introduced a theoretically ideal alternative: ambient concentration contribution permits; see Tietenberg (2006, chap. 4) for an extensive discussion. In the presence of strong spatial heterogeneity, defined by the location of emission points (plants) and receptor points (people), ambient concentration permit markets suffer from practical limitations. Most importantly, such a system requires a large number of markets to work efficiently. The multiplicity of such markets implies high transaction costs and low liquidity of the traded permits.¹ A further problem is the nature of the permit contracts. Unlike emission permits, ambient concentration contribution permits may be difficult to monitor and enforce as additional contributions to emission concentrations at a given receptor location may be subject to large stochastic variation. It is much easier to monitor emissions at the source.

Because of the impracticality of ambient-concentration markets, numerous alternatives have been proposed in the past; see Atkinson and Tietenberg (1982) for a discussion. Much of that research has been focused on second-best models involving trading zones with a limited number of markets. Among the most prominent examples are Førsund and Nævdal (1998), who propose a zonal system with inter-zone exchange rates, and more recently Krysiak and Schweitzer (2010), who derive the optimal size of zonal permit markets in the presence of informational constraints.

Recent research by Farrow et al. (2005), Hung and Shaw (2005), Muller and Mendelsohn (2009), Fowlie and Muller (2013), Fowlie et al. (2015) and Holland and Yates (2015) have opened a new frontier for tackling the hot spot problem; it is

¹ Trading costs, combined with market and regulatory uncertainty, may reduce the efficiency of permit trading systems. Montero (1997) shows that these frictions can have a non-negligible impact on abatement outcomes: trade volume decreases, and total compliance cost increases. In the context of the RECLAIM market in the Los Angeles basin, Gangadharan (2000) finds that a variety of transaction costs reduce the probability of trading in the RECLAIM market in 1995 by 32 %. Transaction costs for permit trading systems consist of a variety of specific costs, which in turn are driven by particular economic factors (Egenhofer 2003). *Search costs*, the cost of matching buyer and seller, are greatly reduced through organized exchanges, but crucially depend on the liquidity and transparency of the market. *Negotiation costs* arise from contracting and standardization of contracts through permit exchanges. These depend on the clarity of the property rights assigned by the contracts. *Monitoring costs* arise through verification of compliance, but are typically borne by the regulator. Similarly, *enforcement costs* arise in the case of non-compliance when the regulator needs to enforce compliance or fine violators. For individual firms there are also *information costs* for monitoring permit markets. Firms may also incur *insurance costs* to allow for the technical risk of accidental non-compliance.

a matter of ‘getting the prices right’ for individual emitters.² Individual emitters need to face heterogeneous price signals that reflect their contribution to environmental damage, and not just their overall emission level. The key innovation in these papers is the use of bilateral trading ratios through which pairs of firms can exchange permits. While the economic logic of this approach is compelling, the question arises on how to operationalize this concept. Obviously, bilateral trading is not as efficient as trading through an integrated (anonymous) permit market with a single price. This paper advances such a solution for ‘getting the market right.’ This is an implementation question—and a policy problem—rather than a question about new economic theory. Theoretical elements introduced in this paper refine and complement the theoretical structure in the aforementioned papers. While some theoretical elements may appear as isomorphic, they introduce subtle modifications and simplifications that help develop an operationally feasible permit market for air pollution hot spots. Therefore, the aim of this paper is to refine *environmental policy*, but not to develop novel *economic theory*.

The keys to a practicable tradeable permit system for ‘hot spots’ are simplicity, objectivity, liquidity, and gradualism. A trading system is simple if it keeps transaction costs low, and offers verifiable and enforceable contracts. To make a permit market for hot spots a feasible alternative to other forms of regulatory intervention, such a market cannot be more complex for participants than conventional permit markets. A market must also strive for objectivity by basing any firm-specific rules on measurable quantities that are relatively immune to challenge or to lobbying. In an UNCTAD report, Tietenberg et al. (1999, pp. 105–107) comments on design principles for a permit trading system:

The emissions trading system should be designed to be as simple as possible. The historic evidence is very clear that simple emissions trading systems work much better than severely constrained ones. The transaction costs associated with implementing and administering an emissions trading system rise with the number of constraints imposed, and as transactions costs rise, the number of trades falls. As the number of trades falls, the cost savings achieved by the programme also decline. [...] Transaction costs play a key role in the success or failure of an emissions trading system. In the past, only emissions trading programmes with low transaction costs have succeeded in substantially lowering the cost of compliance.

To make emission permit trading work for ‘hot spots,’ the trading system has to be as simple and transparent as the conventional cap-and-trade system for uniformly mixed pollutants. As Burtraw (2013) points out, governing institutions are often the pitfall of emission pricing: getting institutional mechanisms ‘right’ is as important as getting the pricing ‘right.’ As the SO₂ trading system has shown, even a simple cap-and-trade system faces formidable obstacles (Schmalensee and Stavins 2013). State-level and source-level constraints—imposed by the *Clean Air Interstate Rule*

² While Farrow et al. (2005) and Hung and Shaw (2005) propose trading ratio systems for water quality, Muller and Mendelsohn (2009) apply trading ratios to air quality. See also Konishi et al. (2015) for an analysis of the pitfalls in the different approaches to trading ratios for water quality.

and related court rulings—effectively closed down the SO₂ market in the United States.

This paper proposes a new practicable emission trading system for non-uniformly mixed pollutants. It implements the pioneering economic theory by Muller and Mendelsohn (2009) through subtle, but useful enhancements. It involves a single emission permit market with permits based on firms' actual emissions. However, to provide the economically efficient incentive for reducing emissions more aggressively in air pollution 'hot spots,' firms contributing emissions to 'hot spots' must face a higher effective permit price than firms outside these 'hot spots.' On the other hand, a unified permit market only has a single market-clearing price. The key insight to facilitate this effective price differentiation is to require firms to buy a number of permits that is *proportional* to their emissions rather than exactly equal to their emissions, with the proportionality factor varying by emission source and reflecting each source's contribution to environmental damage. For example, a firm contributing pollution to a hot spot may have to buy twice as many permits as their actual emissions, while firms outside the hot spot may have to buy permits for only half of their actual emissions. This proportionality measure, henceforth referred to as the *hazard factor*, can be measured objectively and attributed to each firm prior to trading.³ With this adjustment in place, an individual firm faces an effective permit price per unit of emissions that is equal to the product of the market permit price and the hazard factor. Each firm buys emission permits equal to the product of its emissions times its hazard factor at the prevailing permit market price. This feature also allows trading to be anonymous.

The intuition behind this simple concept is that price and quantity signals are both economically efficient (at least without uncertainty), and that both signals can be mixed in a hybrid fashion. Policy hybridization plays an important role in environmental policy discussions since the seminal work by Roberts and Spence (1976) where price floors and price ceilings can improve the overall efficiency of the system by ensuring sufficient dynamic incentives to innovate (price floor) and relief from sudden price spikes (price ceiling, or allowance reserve). A hybrid price-quantity system is envisioned in this paper to overcome limitations of previously proposed emission permit markets for 'hot spots.' While the use of multiple (hybrid) policy instruments has been considered extensively (e.g., Benneer and Stavins 2007) in particular with respect to second-best approaches, the hybrid price-quantity mechanism proposed here aims at achieving first-best optimality, albeit sequentially (iteratively) rather than instantaneously. This hybrid market mechanism is a natural extension of Muller and Mendelsohn (2009) that turns their theoretical instrument into a practical instrument. The proposed hybrid market is also in the spirit of the results in Holland and Yates (2015), who show that under asymmetric information trading ratios based on marginal damage are generally not optimal and can be improved upon through optimal ex-ante trading ratios.

³ The term 'hazard factor' could be understood as a damage premium or damage discount and is the same as the 'trading ratios' in Muller and Mendelsohn (2009). With n firms, arbitrage (consistency) conditions enforce that the $n(n-1)$ bilateral trading ratios are in fact just $n-1$ relative prices.

The trading system introduced in this paper also relies on gradualism: approaching the first-best solution over time rather than in a single step, as is the common approach in most theoretical models. The benefit of gradualism is twofold. It reduces the information burden of determining marginal damage, relying instead on observing average damage changing over time. Gradualism is also a political imperative. Firms must be given sufficient time to make necessary investments and adjustments.

Lastly, the proposed market also allows for *anonymous* trades. Market participants do not need to know the identity of other participants in order to trade permits. Financial markets that facilitate anonymous trading are generally thought to attract more liquidity (Welker 1995; Theissen 2003; Rindi 2008).

2 Model

The purpose of the theory section in this paper is to refine the trading ratios theory developed by Muller and Mendelsohn (2009) and others. The theory presented here is close to but not isomorphic to theirs. It simplifies some elements, enhances others, and in particular allows for robust features that reach the optimal solution even when firms face discrete abatement decisions (and thus have undefined marginal abatement costs) and when local marginal damages cannot be observed and measured directly.

2.1 Regulator's objective

Consider the regulator's problem of controlling the ecosystem hazard⁴ H_j in regions $j = 1, \dots, m$ that emanates from emissions generated by plants $i = 1, \dots, n$. Ecosystem hazard includes the health risk of the population directly exposed to the emissions but may also include the negative impact on animals and plants, which (in addition to their intrinsic value) have an indirect effect on the health of the population through the food chain. Let P_j denote appropriate ecosystem weights ('population') for region j and define regions sufficiently small that emission concentrations can be thought of as homogeneous within that area. Let A_j denote measured ambient pollution concentration in region j in a reference period. As in Hoel and Portier (1994) and in line with Krysiak and Schweitzer (2010), damage follows a quadratic dose–response function.⁵ Goodkind et al. (2014) discuss log-

⁴ I prefer using the term 'hazard' instead of 'damage.' Damage, by definition, describes harm or injury that has occurred already. Hazard, by contrast, describes the *chance* of being injured or harmed; the damage need not have occurred yet. Thus, hazard describes the exposure or vulnerability to harm or injury (i.e., the damage *potential*) and is perhaps a more inclusive term than 'damage.'

⁵ The underlying structure of the model is quadratic in emission concentrations, that is $\sum_j A_j^2$. In a linear model inter-regional trade-offs between regions i and j would be $\partial A_j / \partial A_i = -1$, suggesting that reducing emission concentrations in a 'benign' region i while letting them increase in a 'hot spot' j by the same amount would leave total hazard unchanged. By comparison, with a quadratic function the inter-regional trade-off is $\partial A_j / \partial A_i = -A_i / A_j$. This means that to keep total hazard unchanged, an increase in ambient concentrations in 'benign' region i would only offset a smaller ambient concentration increase in 'hot

linear and log-log dose response functions. The ecosystem damage (hazard) in region j can be defined as

$$H_j = P_j A_j \left[A_j^\square + \omega \sum_{i=1}^n d_{ij} E_i \right] \approx P_j A_j^2 \quad (1)$$

where E_i is the emission from plant i and d_{ij} is the dispersion factor between firm i and location j (the fraction of firm i 's emissions deposited in region j) such that $\sum_j d_{ij} = 1$. A_j^\square captures ambient concentrations from natural or transboundary sources, and ω converts emission deposits into emission concentrations. Even though the hazard function is quadratic in emission concentrations, including P_j makes this more general. It is a region-specific “fixed effect” that can capture any type of conditioning factor that influences environmental hazard.

An important innovation in this model is that the non-linear dose–response function is captured through the product of an ex-ante (observed) emission concentration A_j and an ex-post (incentivized) emission concentration captured by the term in square brackets; thus H_j may be thought of as $P_{j,t-1} A_{j,t-1} A_{j,t}$ for successive time periods t . Applying this temporal logic allows for a loss function quadratic in A_j while keeping the model effectively linear; see “Appendix 1” for mathematical details. This novel technique entails that the welfare loss is minimized only over time, not instantaneously. Convergence to the minimum requires that $\sum_j |A_{j,t} - A_{j,t-1}|$ remains sufficiently small. Using temporal logic to linearize the economic decision problem comes at little economic cost because a regulator will phase in a cap-and-trade system gradually to allow firms to adjust. Iterating this second-best model over several periods assures convergence to the first-best outcome. The regulator decreases the emission quota gradually until the welfare loss reaches its observed minimum. Unlike static models that need to achieve optimality in one shot, the model introduced here gets there too, but in smaller steps.

Equation (1) recognizes that the total ecosystem damage (hazard) is not merely a function of ambient concentrations, but also a function of the number of people, plants, and animals that are exposed to the concentration. This implies a trade-off between higher ambient concentrations in sparsely populated areas against lower concentrations in densely populated areas. Different regulators may therefore rely on different weights P_j . For example, P_j could simply denote population, or it could capture pollution-related health care costs. The US Environmental Protection Agency recognizes the importance of this concept in its calculation of *Risk-Screening Environmental Indicators* based on data from its *Toxics Release Inventory*.

The regulator is also concerned about the (expected) abatement cost $B = \sum_i B_i$ incurred by firms in meeting the desired environmental footprint. Abatement costs

Footnote 5 continued

spot” j to leave total hazard unchanged. As j is a “hot spot” and thus $A_i < A_j$, the ratio $A_i/A_j < 1$. Earlier work on emission hot spots often posits a constrained optimization problem so that $\forall j: A_j \leq \bar{A}$. This set-up rules out any inter-regional trade-offs and implies $\partial A_j / \partial A_i = \infty$, a very strong Rawlsian view of the world. Contrastingly, the weighted quadratic loss function used here reflects a utilitarian approach.

are observed by the firm but are unobservable to the regulator. Following Baumol and Oates (1988), the regulator’s objective is to minimize the ‘welfare loss’ L from ecosystem hazard H and abatement cost B

$$L = \sum_{i=1}^n B_i + \gamma \sum_{j=1}^m H_j \tag{2}$$

Here, γ represents the marginal damage of a unit of ecosystem hazard. Extending Baumol and Oates (1988), γ may also be thought of as a political preference parameter indicating the regulator’s weight on ecosystem hazard. Importantly, hazards are additive over regions, allowing inter-regional hazard trade-offs.

The regulator issues permits equal to the targeted total amount of emissions $\bar{E} \equiv \sum_i E_i$. The regulator assigns hazard factors ϕ_i to individual firms. Market participants then determine the market price τ for these permits. Individual firms buy $\phi_i E_i$ permits, and therefore $\bar{E} = \sum_i \phi_i E_i$ must hold as well.⁶

2.2 The plant

Each plant i ’s emissions affect the neighboring regions j through the dispersion factor d_{ij} so that $\sum_j d_{ij} = 1$. These dispersion factors are determined by measurement or modelling of atmospheric dispersion processes. Each plant has final emissions E_i and is required to purchase $\phi_i E_i$ emission permits at market price τ , where $\phi_i > 0$ is each plant’s hazard factor as determined by the environmental regulator at the beginning of each trading period. Consequently, firm i faces an effective unit price for its emissions E_i equal to $\phi_i \tau$.

Plant i ’s final emissions are a function of its abatement effort $\theta_i \in [0, 1[$ so that $E_i = (1 - \theta_i)E_i^0$, with E_i^0 denoting the original unabated emission level of plant i . The scale (output) of the firm is assumed to be fixed.⁷ Given an abatement cost function $B_i(\theta_i)$, the firm’s objective is to minimize its environment-related costs

$$C_i = \min_{\theta_i} \{B_i + \tau \phi_i E_i\} \tag{3}$$

by choosing the optimal abatement effort θ_i in response to the effective price $\tau \phi_i$ for buying permits for its (post-abatement) emissions E_i . Straight-forward minimization yields optimal abatement effort θ_i^* and corresponding post-abatement levels of emissions E_i^* and abatement cost B_i^* (star superscripts indicate the optimum solution.) After abatement, the firm buys $\phi_i E_i^*$ permits at price τ .

Economic theory prefers the use of general forms over specific functional forms as the latter may generate results that only hold under that particular choice of

⁶ The emission permit market with hazard factors ϕ_i entails two limiting cases. First, setting $\phi_i = 1$ replicates a conventional cap-and-trade market. This is a useful benchmark against which one can compare the benefits of introducing plant-specific hazard factors. This limiting case is also approached when the number of regions drops to $m = 1$, which implies that the world is a single ‘hot spot.’

⁷ This simplification rules out quantity adjustment in response to environmental price signals. This is a useful and plausible approximation when abatement costs play only a small part in the overall cost structure of a firm.

functional form. On the other hand, specific functional forms are necessary to explore solutions numerically. In the context of abatement costs, specific functional forms can also be used to clearly distinguish between an *intensive margin* and an *extensive margin* of abatement activity. Abatement activity varies continuously (and monotonically) along the intensive margin. The extensive margin of pollution abatement identifies firms commencing or ceasing abatement activity, and thus identifies threshold effects. Importantly, many firms engage only in discrete abatement decisions at the extensive margin, such as installing a scrubber or electrostatic precipitator.

In what follows three functional forms for abatement costs are explored simultaneously: a logarithmic form, a quadratic form, and a stepwise form. Together they cover a wide—and economically plausible—range of applications. Table 1 defines successively: the total abatement cost B_i for plant i ; the marginal abatement cost $dB_i(\theta_i)/d\theta_i$; the marginal abatement cost elasticity $(d^2B_i/d\theta_i^2)/(dB_i/d\theta_i)\theta_i$; the optimal abatement effort θ_i^* ; the extensive margin for commencing abatement $\theta_i^* > 0$; the extensive margin for completing abatement $\theta_i^* < 1$; the post-abatement emission level $E_i^*(\theta_i^*)$; and the post-abatement abatement cost $B_i^*(\theta_i^*)$. These three functional forms are purposefully simple and allow the derivation of concrete and insightful algebraic and numeric solutions. While the generality of the hazard factor model has already been established in Muller and Mendelsohn (2009) by way of ‘trading ratios,’ the specificity of the functional forms provides new insights especially about the differences between the intensive and extensive margins of abatement activity.

All three functional forms have a single cost factor b_i that captures the heterogeneity across firms. The crucial information asymmetry between regulator and firm is that the regulator cannot observe b_i ; the firm knows its true b_i but not the regulator. The functional forms also have several desirable algebraic properties. Both logarithmic and quadratic form are convex; increasing abatement effort raises abatement costs. In the quadratic case the abatement cost elasticity is exactly one,

Table 1 Specific abatement cost functions

	Logarithmic	Quadratic	Stepwise
B_i	$-b_i \ln(1 - \theta_i)E_i^0$	$\frac{b_i}{2} \theta_i^2 E_i^0$	$1(\tau\phi_i > b_i)(b_i E_i^0)$
$dB_i(\theta_i)/d\theta_i$	$\frac{b_i}{1 - \theta_i} E_i^0$	$b_i \theta_i E_i^0$	Step function
$\frac{d^2 B_i/d\theta_i^2}{dB_i/d\theta_i} \theta_i$	$\frac{\theta_i}{1 - \theta_i}$	1	–
θ_i^*	$1 - \frac{b_i}{\tau\phi_i}$	$\frac{\tau\phi_i}{b_i}$	0 or 1
$\theta_i^* > 0$?	$\tau\phi_i > b_i$	Always	$\tau\phi_i > b_i$
$\theta_i^* < 1$?	Always	$\tau\phi_i < b_i$	$\tau\phi_i < b_i$
$E_i^*(\theta_i^*)$	$\frac{b_i}{\tau\phi_i} E_i^0$	$\left(1 - \frac{\tau\phi_i}{b_i}\right) E_i^0$	$1(\tau\phi_i > b_i) E_i^0$
$B_i^*(\theta_i^*)$	$b_i \ln\left(\frac{\tau\phi_i}{b_i}\right) E_i^0$	$\frac{(\tau\phi_i)^2}{2b_i} E_i^0$	$1(\tau\phi_i > b_i)(b_i E_i^0)$

while in the logarithmic case it increases from zero to infinity as abatement effort approaches one. Importantly, the logarithmic form has an extensive margin for commencing abatement (the permit price has to be high enough to trigger abatement), while the quadratic form has an extensive margin for ceasing abatement (the permit price reaches a level at which the firm abates all its emissions).

The stepwise function only has an extensive margin but no intensive margin. A firm deploys abatement if the permit price reaches a critical level ($\tau\phi_i = b_i$), and when it does so it abates all of its emissions. This is not an unrealistic assumption for individual firms that often have only a single technological option such as switching inputs, production methods, or deploying an abatement device. Abatement effort is thus either $\theta_i = 0$ or $\theta_i = 1$, and emissions are either zero or E_i^0 . The ranking of b_i establishes an ‘abatement ladder’ along the lines of Antweiler (2003). Lastly, $1(\cdot)$ denotes the indicator function.

2.3 Optimal intervention

With the above results the regulator’s consolidated objective function for the logarithmic abatement cost function becomes

$$L = \sum_i b_i E_i^0 \ln\left(\frac{\tau\phi_i}{b_i}\right) + \gamma \sum_j P_j A_j \left[A_j^\square + \omega \sum_i d_{ij} \frac{b_i E_i^0}{\tau\phi_i} \right] \tag{4}$$

It is analogous for the quadratic case. For the stepwise abatement cost function the objective function is simply

$$L = \tau \sum_{\Theta_0} \phi_i E_i^0 + \sum_{\Theta_1} b_i E_i^0 + \gamma \sum_j P_j A_j \left[A_j^\square + \omega \sum_{\Theta_0} d_{ij} E_i^0 \right] \tag{5}$$

where the set $\Theta_0 \equiv \{i | \tau\phi_i \leq b_i\}$ captures non-abating firms and the set $\Theta_1 \equiv \{i | \tau\phi_i > b_i\}$ captures all fully abating firms.

The regulator’s problem is finding the optimal intervention that minimizes (4) or (5). Establishing hazard factors ϕ_i with certainty then solves the ‘hot spot’ problem implicitly. The optimal intervention involves minimizing L with respect to an individual price for each plant, $\tau_i \equiv \tau\phi_i$. In the case of the logarithmic and quadratic abatement cost function, this optimization yields prices

$$\tau_i^* = \gamma\omega \sum_j P_j A_j d_{ij} = \gamma Q_i \tag{6}$$

where $Q_i \equiv \omega \sum_j P_j A_j d_{ij}$ are hazard weights for plant i , and thus the plant-specific hazard contribution is $H_i \equiv Q_i E_i$. The next step is demonstrating that finding a single permit price τ replicates the first-best intervention (6). The first-order condition for minimizing (4) with respect to a single permit price τ , given hazard factors ϕ_i , yields

$$\tau^* = \gamma \frac{\sum_i H_i}{\sum_i E_i \phi_i} = \gamma \frac{H}{E} \quad (7)$$

where $\bar{Q} \equiv H/E = \sum_i Q_i E_i / \sum_i E_i$. Now consider the solution

$$\phi_i = \frac{H_i/H}{E_i/E} = \frac{Q_i}{\bar{Q}} \quad (8)$$

which defines the hazard factor as the ratio of hazard share to actual (post-abatement) emission share. From this it follows immediately that $\tau^* = \gamma$, and thus $\tau_i^*/\tau^* = Q_i/\bar{Q}$. As long as the regulator is able to determine $\phi_i = Q_i/\bar{Q}$ reliably, the enhanced version of the permit market is able to obtain the first-best solution for emission reduction. The simple modification that leads to this crucial result is to require firms to buy $\phi_i E_i$ permits instead of merely E_i as in a conventional cap-and-trade market. Thus, the ‘cap’ is effectively on hazard despite being nominally on emissions (\bar{E}).

Now consider the case of the stepwise abatement cost function as captured in (5) and again assume (8) as the optimal hazard factor. Then the indicator function defines the set of abating firms as those where $\tau \phi_i = \gamma H_i/E_i > b_i$. Put in words, firms whose hazard-to-emission ratio, normalized against the average hazard-to-emission ratio, exceeds the abatement cost factor b_i/τ are selected into the set Θ_1 of fully abating firms. This selects firms which are either particularly hazardous (high H_i/E_i) or are particularly low-cost abaters (low b_i). The ratio $b_i/(H_i/E_i)$ ranks all firms along a single dimension. Without loss of generality, assume that firms are ranked from lowest to highest so that low i end up in the non-abater group Θ_0 and high i end up in the full-abater group Θ_1 .

To prove that (7) and (8) minimizes (5), consider the marginal firm (k) that is indifferent between abating and non-abating. For this firm, it must hold that $\tau \phi_k = b_k$. Now construct two versions of (5), L^+ with firm k abating, and L^- with firm k non-abating. Using (7) and (8), $\tau_i \phi_i E_i^0 = \gamma (H_i/E_i) E_i = \gamma H_i$. Thus $\sum_{\Theta_0} \tau \phi_i E_i^0 = \gamma \sum_{\Theta_0} H_i$, and similarly it follows that the rightmost term in (5) is $\gamma \sum_{\Theta_1} H_i$. With $\sum_{\Theta_0} H_i + \sum_{\Theta_1} H_i = H$, it follows that $L^+ = \gamma H^+ + \sum_{\Theta_0^-} b_i E_i$ and $L^- = \gamma H^- + \sum_{\Theta_0^-} b_i E_i$. For the marginal firm it must hold that $L^+ = L^-$, and therefore $\gamma(H^+ - H^-) = b_k E_k$. By construction, $H^+ - H^- = H_k$, the hazard from firm k . Hence, the regulator’s no-arbitrage conditions requires $\gamma(H_k/E_k) = b_k$. It is immediately apparent that this condition is identical to the firm’s indifference condition because $\tau_k \phi_k = \gamma(H_k/E_k)$. This concludes the proof.

The above has demonstrated that the regulator’s choice of ϕ_i through the ratio (8) holds under a variety of functional forms, allowing both for intensive and extensive margins of abatement. This simple rule of setting ϕ_i based on (normalized) hazard-to-emission ratios is economically intuitive, robust across a variety of specifications, and can be established plausibly and accurately using data from the previous trading period.

2.4 Relationship to trading ratios

Using a general environmental damage function D , Muller and Mendelsohn (2009) derive an optimal set of trading ratios TR_{ij} for a pair of firms i and j so that

$$TR_{ij} = \frac{\Delta E_i}{\Delta E_j} = \frac{\partial D_j / \partial E_j}{\partial D_i / \partial E_i} \tag{9}$$

is the ratio of marginal damage from firm j to the marginal damage from firm i . Implicitly, each firm has a price P_i at which it can sell permits to other firms. and assuming a competitive market for each firm’s permits, the price ratio of the two permits must equal the trading ratio.

The problem with bilateral trading ratios is that it does not fully solve the multiple market dilemma encountered before. Whereas Montgomery (1972) established markets for each of m regions, Muller and Mendelsohn (2009) establishes markets for n firms. There is no central market through which permits can be bought and sold at a single price. By comparison, the market model introduced here solves this problem by assigning adjustment factors ϕ_i for each firm through which they can trade permits in a single integrated market at a single price τ . The trading parties are anonymous as no bilateral bargaining or contracting needs to take place.

How does the market introduced here differ from Muller and Mendelsohn (2009)? Here, the equivalent to their trading ratios is the ratio of hazard factors. Thus,

$$TR_{ij} = \frac{\phi_j}{\phi_i} = \frac{H_j/E_j}{H_i/E_i} \tag{10}$$

captures the ratio of average hazard for plant j relative to the average hazard for plant i . In other words, the relative *marginal* damage concept (in a single time period) has been replaced by a relative *average* damage concept—but only over time. As a first-best solution, the use of marginal damages gets the prices right in a ‘single shot.’ Using average damages gets prices right only gradually as the market iterates towards the welfare optimum. Why would regulators prefer using average

Table 2 Numerical illustration: parameters

Firm			Region $j(d_{ij})$				\sum_j	χ_i
i	b_i	E_i^0	a	b	c	d		
1	10	150	0.7	0.2	0.1	0.0	1.0	1.656
2	7	70	0.6	0.3	0.0	0.1	1.0	1.500
3	5	100	0.5	0.1	0.1	0.5	1.0	1.259
4	8	130	0.1	0.0	0.3	0.6	1.0	0.370
5	9	250	0.0	0.1	0.2	0.7	1.0	0.215
Weight P_j			24	18	8	5	30	
Initial A_j			210	86	114	310	–	
Initial γH_j			5292	666	520	2403	8880	

damages over marginal damages? After all, the APEEP model developed by Muller and Mendelsohn (2009) provides detailed estimates of marginal damages in the United States. However, their model relies on functional form to identify marginal damage (Muller and Mendelsohn 2009, p. 1734, eq. 34). Rather than imposing specific functional forms, an iterative procedure that relies on trading ratios defined by average damage reaches the same optimum. Average damage is readily observable and thus reduces regulatory complexity.

3 Numerical illustration

A simple numerical simulation can help illustrate the key concepts in this paper. Using the logarithmic abatement cost function, Table 2 provides a ‘sandbox’ for exploring the model with five firms (1–5) and four regions (a–d). Region (a) is

Table 3 Numerical illustration: simulation

τ	B	H	L	E	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
0.00	0	8880	8880	700.0	1.379	1.279	1.243	0.599	0.499
5.00	179.1	8346	8525	669.9	1.549	1.436	1.396	0.673	0.561
6.00	378.1	7022	7400	644.4	1.563	1.454	1.431	0.716	0.607
7.00	659.4	6131	6791	615.7	1.552	1.448	1.448	0.752	0.648
8.00	967.4	5346	6313	588.2	1.529	1.432	1.460	0.792	0.696
9.00	1228	4806	6034	567.6	1.501	1.411	1.465	0.824	0.735
10.00	1516	4312	5828	543.9	1.475	1.393	1.464	0.846	0.763
11.00	1842	3835	5677	517.1	1.454	1.378	1.461	0.861	0.785
12.00	2266	3311	5577	481.7	1.446	1.372	1.460	0.865	0.792
12.85	2662	2887	5549	449.8	1.446	1.372	1.460	0.865	0.792
Δ (%)		–67.5	–37.5	–35.7	4.81	7.27	17.46	44.38	58.62

τ	A_a	A_b	A_c	A_d	θ_1	θ_2	θ_3	θ_4	θ_5
0.00	210.0	86.00	114.0	310.0	0	0	0	0	0
5.00	194.8	82.64	111.2	295.6	0	0.025	0.283	0	0
6.00	180.8	77.68	109.8	287.7	0	0.198	0.418	0	0
7.00	163.3	72.04	107.7	282.5	0.080	0.309	0.507	0	0
8.00	145.9	66.65	105.5	278.7	0.182	0.389	0.572	0	0
9.00	132.8	62.58	103.9	275.8	0.260	0.449	0.621	0	0
10.00	121.6	59.30	100.5	269.3	0.322	0.497	0.659	0.055	0
11.00	111.6	56.57	95.45	259.7	0.375	0.538	0.689	0.155	0
12.00	102.7	52.75	88.89	243.1	0.424	0.575	0.715	0.230	0.053
12.85	95.88	49.26	83.01	227.0	0.462	0.603	0.734	0.281	0.116
Δ (%)	–54.3	–42.7	–27.2	–26.8					

densely populated and industrial, region (b) more suburban, region (c) large and rural, and region (d) a small industrial town. Parameter values for abatement costs b_i and initial emissions E_i^0 are given along with dispersion factors d_{ij} and regional weights P_j . The table also shows the initial ambient concentrations (prior to abatement) A_j and the corresponding hazard γH_j , expressed in monetary terms using a hazard cost factor of $\gamma = 1/200$. Regions (a) and (d) are ‘hot spots’ as their initial ambient emission concentration is two or three times that of the neighbouring regions (b) and (c), and thus the regulator’s assessment of total hazard of regions (a) and (d) is about five times that of regions (b) and (c). The spatial interconnections among the five firms can be expressed through the centrality measure χ_i (see ‘Appendix 2’ for definition); firms 1–3 have above-average centrality and firms 4 and 5 have below-average centrality.

Table 3 shows the results of a market simulation where the number of emission permits is gradually lowered from 700 until the regulator’s loss function reaches a minimum. The first row shows the baseline (zero market price τ) and intermediate steps are shown for ‘round’ permit prices. The final permit price turns out to be $\tau = 12.85$, and the last row Δ (%) shows percentage changes of the optimal result against the baseline.

While total emissions E are reduced by 36 %, total environmental hazard is reduced by about 68 %. Columns ϕ_1 – ϕ_5 show how the hazard factors evolve. In part, the changes are driven by neighbouring plants commencing abatement activity as the permit price reaches their participation constraint $\tau\phi_i > b_i$, which is visible in the bottom right segment of the table for abatement effort θ_1 – θ_5 . Plant 3 has the lowest abatement cost and eventually reduces 73 % of its emissions, whereas other plants only reach an abatement effort of between 12 and 60 %. The hazard factors only evolve slowly.

Ambient emission concentrations in the four regions drop between 27 % (regions c and d) and 54 % (region a). This illustrates that the market helps reduce hazard more in the more populous regions (a) and (b), whereas region (d) with the highest original emission concentration has much smaller population, and consequently sees a lesser reduction. Of course, it also matters whether plants close to this region have affordable abatement ability. In the example, plants 4 and 5 contribute much to region (d) but have relatively high abatement costs.

Table 4 provides a comparison to a conventional cap-and-trade market in which $\phi_i = 1$. At its optimum level, the permit price is higher (14.25 versus 12.85) for a larger reduction in emissions (42 vs. 36 %) but about the same reduction in overall hazard. Comparing total losses (33.3 vs. 37.5 %), the conventional cap-and-trade

Table 4 Comparison with conventional cap-and-trade

τ	B	H	L	E	θ_1	θ_2	θ_3	θ_4	θ_5
14.25	3038	2887	5925	405.6	0.298	0.509	0.649	0.439	0.368
Δ (%)		–67.5	–33.3	–42.1					

market underperforms as expected, and generates a comparable level of hazard reduction at 15 % higher cost.

4 Practical considerations

One of the key objectives of this paper is to demonstrate the practical feasibility of the proposed hybrid price-quantity emission permit market. It is therefore useful to discuss a number of potential issues and possible extensions.

4.1 Multiple pollutants

An emission permit market has the potential to target multiple pollutants rather than just one. For example, a metallic air toxics market could include a list of substances with similar characteristics. Including multiple pollutants in a single market can be very desirable when pollutants are substitutes for each other because regulating only one pollutant may cause undesirable shifting of production methods towards unregulated pollutants. An emission permit system can be designed to cover multiple pollutants by assigning fixed toxicity factors to each pollutant.⁸ For example, the US-EPA (2004) Risk-Screening Environmental Indicators (RSEI) of Chronic Human Health provide suitable equivalence factors.

4.2 Bioaccumulation and biomagnification

Many toxics have the tendency to bioaccumulate and biomagnify. For example, lead, mercury and cadmium emissions from a smelter in a sparsely populated area may expose a much larger population far away if these air toxics find their way into the food supply through nearby agricultural production or fish habitat. It is possible to design ecosystem weights P_j for each region that allow for health hazards ‘imported’ from other regions. However, spatial models that take these types of second-order effects into account have not been widely developed yet. Nevertheless, the emission permit market envisioned in this paper is able to deal with such extensions because the regulator can calculate appropriate P_j , which are then imputed into specific hazard factors ϕ_i for individual firms.

4.3 Transboundary pollution

Air pollutants travel easily across national boundaries. Yap et al. (2005), a study commissioned by the Ontario Ministry of the Environment, concluded that of the estimated \$9.6 billion in health and environmental damages from ground-level ozone and fine particulate matter in Ontario in 2003, 55 % is attributable to U.S. emissions. Transboundary pollution, which in the model is captured through A_j^\square , is a spatial fixed effect that does not impact the design of the permit market. The term

⁸ The practice of equivalency factors is well established for greenhouse gases, where emissions are expressed in carbon dioxide equivalency units.

drops out in the first-order conditions of the model. It only matters insofar as the regulator sets more aggressive emission reduction goals in order to compensate for the transboundary pollution, in which case local firms bear the burden of abatement that more efficiently could be borne by firms across the border.

4.4 Permit banking and borrowing

Emissions of air pollutants are often more variable over time than fundamental indicators of operational size, such as plant employment or value added. The spatial ‘hot spot’ problem may thus become aggravated through temporal clustering. Permit banking and borrowing are methods to address temporal variability in emissions; see Rubin (1996), Cronshaw and Kruse (1996) and Schennach (2000). A permit system that is defined in terms of ambient concentrations cannot easily accommodate banking or borrowing because of potential temporal clustering of emissions. Temporal clustering would intensify ‘hot spots’ of very high peak concentrations, posing a significant health risk. The economic rationale for banking and borrowing is compelling: greater flexibility across time provides for a better allocation of abatement investments over time. Allowing consolidation of permits over a longer time horizon may contribute to intertemporal efficiency in the presence of high emission variability across years. Newell et al. (2005) explore a bankable permit system in a multi-period setting and demonstrate that it can provide similar outcomes as a price-based system. They explore several approaches towards hybridization, and it is apparent that this flexibility would also be needed in a hybrid market for air pollution hot spots. Permit banking or borrowing may become problematic when firms anticipate changes of their hazard factors ϕ_i over time and use banking or borrowing strategically. More important, however, is the need to establish a permit bank as explored in Akao and Managi (2013). They show that a tradable permit system may not achieve efficiency without setting appropriate permit interest rates, and that such permit interest rates can be generated endogenously without government intervention.

4.5 Plant entry, exit, and relocation

The level of strategic interaction among firms depends on the level of ‘pollution-plant centrality,’ an empirical concept described in greater detail in “Appendix 2”. In essence, the closer plants are located to each other, the more they care about each others’ actions. Entry of a new plant creates a negative externality on surrounding plants, while exit creates a corresponding positive externality. New entrants can aggravate the ‘hot spot’ problem through increased emissions. Concretely, a new entrant will contribute more emissions and larger hazard, which will be reflected in the reference period in an increase in A_j . This in turn increases the hazard factor ϕ_i that the regulator announces for firm i in the next period. The total effect is inversely related to firm centrality. An effective permit system may also help bring about beneficial relocation of firms when relocation is cheaper than abating emissions or buying permits. While relocating a plant does not reduce emissions, it can

significantly reduce environmental hazard if the plant is moved out of a pollution ‘hot spot.’

4.6 Constrained hazard factors

Hazard factors may stretch over a wide range empirically. Firms that face a high ϕ_i would bear a large share of the financial burden of pollution abatement. This may not always be politically feasible. The hazard factors can be suitably constrained to an interval $[\underline{\phi}, \bar{\phi}]$ that the regulator finds politically expedient with respect to distributional equity considerations. In particular, capping ϕ_i at a particular $\bar{\phi}$ may help convince participating firms that the abatement burden is not distributed too unequally and thus may safeguard against distortions of their competitive position. Fowlie et al. (2010) point to the links between initial permit allocation, market structure, and entry and exit outcomes. A permit market with constrained hazard factors diminishes its cost-effectiveness with respect to lowering environmental hazard, but this outcome may still be preferable if the alternative is either no regulation or economically inefficient command-and-control regulation. Constraining hazard factors may also mitigate market structure effects. As Lee et al. (2013) show, market structure effects are even more critical in the context of international emission permits where there are repercussions from trade. Emission markets may be imperfectly competitive, with firms in one country exercising more market power than those in another. The same logic also applies to markets for air pollution hot zones that cross jurisdictions. Emission trading across jurisdiction boundaries may entice local governments to engage in strategic behavior (e.g., subsidizing or taxing permit trading) to shift costs outside their own jurisdiction, which could undermine the efficiency of permit trading. Constraining hazard factors may be necessary to dampen repercussions from endogenous policy responses, as constraints reduce the potential for large differences of economic burden across jurisdictions.

5 Potential applications

The market structure for issuing hybrid permits provides a great deal of flexibility to accommodate particular political constraints on implementation. The sulfur dioxide (SO₂) market in the United States is a key example for the potential as well as the obstacles for implementing a market based on differentiated prices and hazard factors.

The acid rain program in the United States is often acknowledged as the first large-scale attempt to use emission permit trading to pursue reduction of a regional pollutant, sulfur dioxide (Siikamäki et al. 2012; Schmalensee and Stavins 2012). Launched in 1995, it is considered a success as it helped to cut emissions in half. However, permit prices were volatile; they peaked at over \$1,600 per ton of SO₂ in late 2005, but collapsed in 2006 and dropped to virtually zero by the end of 2010. The collapse was primarily a result of an expansion of the reach of the program to more than two dozen Eastern states and a subsequent change in rules. A court challenge brought against the expansion ended in a 2008 ruling by a US Court of

Appeals that allowed the expansion to proceed but required the Environmental Protection Agency to modify the trading program.⁹

In 2005, the *Clean Air Interstate Rule* (CAIR) adopted by the EPA required point sources within non-attainment states to surrender two additional allowances for every ton of SO₂ emission, effectively reducing the cap in those states significantly. Note that this system resembles the model proposed in this paper to some degree: it amounts to hazard factors that are equal across all plants in one state. As firms anticipated the tougher restrictions under CAIR, prices spiked due to the ability to bank permits.

The court ruling in 2008 prompted the EPA to develop a new rule, the *Cross-State Air Pollution Rule* (CSAPR), which was adopted in 2011.¹⁰ It allows for intrastate trading but limits interstate trading to two separate groups of states. Plants in states with binding CSAPR caps are essentially forced to implement abatement technologies and do not have the option of buying permits from plants outside the state that could reduce emissions more cheaply.¹¹ The tight restrictions on interstate trading destroyed the SO₂ market in the United States. CSAPR was implemented fully in January 2015, and acid rain allowances are now traded for less than \$1 per ton. CASPR effectively replaces the acid rain allowances with four separate markets for annual NO_x, seasonal (summertime) NO_x, and SO₂ groups 1 and 2.¹² The stricter CASPR targets are ambitious and will undoubtedly improve health outcomes. However, could the same improved health outcomes be achieved at a lower cost?

The type of market proposed in this paper could have addressed the cross-state challenges effectively and provided improvements in ambient emission concentrations more cost-effectively by targeting the power plants with the highest hazard contributions. Under CSAPR, regulators have to calculate state-level caps so that there is an expectation that downwind states will be able to meet *National Ambient Air Quality Standards* (NAAQS) targets. Thus, regulators have to conjecture required state-level emission reductions and may impose caps that are too strict—triggering court challenges in turn.¹³

⁹ There are two additional contributing factors for the drop in prices. On the market side, demand for coal being was replaced increasingly by demand for inexpensive natural gas. Furthermore, induced by the high emission prices, more and more power plants have installed flue gas desulfurization and selective catalytic reduction equipment to reduce SO₂ and NO_x.

¹⁰ The US *Clean Air Act* contains a “good neighbour” provision under which each *State Implementation Plan* (SIP) must prohibit emissions that will significantly contribute to non-attainment of *National Ambient Air Quality Standards* (NAAQS) in a downwind state.

¹¹ Technically, the Environmental Protection Agency (EPA) sets a pollution cap in each states covered by CSAPR. Power plants are bound by the state-wide cap, allowing for unlimited intra-state trading. Inter-state permit trading is permitted but cannot be used to meet state-level emission caps. See Federal Register 77(113), pp. 34830–34846, June 12, 2012 (<http://www.gpo.gov/fdsys/pkg/FR-2012-06-12/pdf/2012-14251>).

¹² Group 1 sources can only trade with group 1 sources; group 2 sources can only trade with group 2 sources. Group 1 contains 16 northeastern and central states, and group 2 contains 6 mostly southeastern states.

¹³ In July 2015 the US Court of Appeals for the District Court of Columbia Circuit ruled that the EPA must reconsider the 2014 Budgets for SO₂ and Ozone NO_x in several states because they required the states to reduce emissions beyond the points necessary to achieve air quality improvements in downwind areas.

The key problem that led to the adoption of the CSAPR is downwind pollution across state lines. Downwind effects can be modelled effectively, and in the proposed hybrid market in this paper are captured into plant-specific hazard factors. Upwind firms end up with a higher hazard factor—and thus face higher incentives to abate emissions. However, if there are several such firms in proximity to each other, they can trade permit among them in order to reduce emissions cost-effectively. At the same time, plants in downwind states with high ambient concentrations are not penalized excessively for imported transboundary pollution; it is only their own emission plume that matters. Given the spatial extent of many US states, location within a state matters.

The hybrid market envisioned in this paper would be a novel way to implement the economically efficient pricing principles pioneered by Muller and Mendelsohn (2009) and others. “Getting the market right” in the context of the acid rain market in the United States would essentially return a much-improved version of the CAIR: instead of using coarse state-level hazard factors, the proposed market would introduce much more precise plant-level hazard factors (ϕ_i). Increased hazard in some areas due to higher population density can be acknowledged by using location-specific weights (P_j) in the hazard function. Alternatively, locations with NAAQS non-attainment could be penalized with higher weights P_j based on relative morbidity, the ratio of local morbidity to base-level morbidity. Power plants with high hazard factors would find it cheaper to invest in pollution abatement equipment or reduce output than to purchase relatively pollution permits that are made expensive by a high ϕ_i .

The state-level caps imposed by CSAPR leave little flexibility to be gained through emission permit trading and instead rely on “getting state-level targeting right.” Contrastingly, the proposed market concept is more precise because it targets individual plants rather than individual states. Whether any market solution with unrestricted interstate trading is compatible with the 2008 court ruling remains to be seen—the EPA never tried because setting state-level targets was the most direct path to complying with the court ruling. The complexity of the new CSAPR trading system is likely to pose new challenges that perhaps open the door again to reforms in the future.

The CSAPR trading system also introduces a dedicated permit market to deal with seasonally high levels of NO_x concentrations, in addition to a conventional annual emission market for NO_x . Such temporal clustering constitutes a challenge for conventional permit markets, which typically have accumulation periods of one or more years. The trading system proposed in this paper has sufficient flexibility to allow for temporal clustering and thus obviates the need for two separate emission permit markets. Recall that plant-level hazard contributions are defined as $H_i \equiv \omega E_i \sum_j P_j A_j d_{ij}$. If all ambient concentrations A_j varied proportionally across seasons, this would not affect the plant-specific hazard factors ϕ_i . However, if seasonal concentrations are region specific, there will be different hazard factors in different seasons. In practice, plants that affect locations with high seasonal pollution concentrations are forced to buy more permits during the high season than the low season for the same amount of actual emissions. For example, a power plant

with $\phi_i = 0.80$ in the winter may have $\phi_i = 1.25$ in the summer, which would provide a strong incentive to reduce output in the summer (or install additional abatement equipment). The beauty of the hybrid trading system is that a single unit of a hybrid emission permit can mean different physical units of emissions for *different plants* as well as *different seasons*. Hybridization introduces a significant degree of flexibility, both spatially and temporally, that is absent in conventional trading systems.

6 Conclusion

Limited geographic dispersion of air pollutants creates ‘hot spots’ of environmental damage (or more generally, environmental hazard), accentuated by varying population density and ecosystem characteristics. Such ‘hot spots’ pose a formidable challenge for market-based instruments. High transaction costs make it infeasible to operate a large number of regional ambient concentration contribution permit markets as envisioned in the pioneering work of Montgomery (1972). Recent work by Farrow et al. (2005) and Muller and Mendelsohn (2009) has introduced trading ratios as a solution to the ‘hot spot’ problem, further extended by Holland and Yates (2015). Their work succeeds in ‘getting the prices right’ for the hot spot problem. Taking their work one step further is ‘getting the market right:’ designing a single integrated emission permit market with low transaction costs as well as regulatory and operational simplicity.

The permit market proposed in this paper has the desirable property of iterating over time towards the optimal economic outcome as the regulator tightens the emission cap gradually. Rather than using firm-pair ‘trading ratios,’ each firm is assigned a single ‘hazard factor’ by the regulator prior to each trading period. This factor is based on observables established during the previous trading period. Firms buy permits in proportion to their emissions *and* their specific hazard factors. This amounts to a hybrid price-quantity instrument in which firms face an effective price per unit of emission that varies along with their specific hazard factor, while on a practical level they purchase and trade permits in a (liquid) market with a single price. This novel hybrid policy instrument implements the general theoretical solutions established in the literature. The simplicity of the proposed market, and the related reduction in transaction costs, makes it particularly appealing from a policy perspective.

Perhaps the key insight in this paper is that permit markets can be allowed to iterate towards the optimal solution, rather than moving to the optimal efficient solution in a single step as envisioned in most of the literature. The notable exception is Goodkind et al. (2014), who demonstrate the value of an iterative approach convincingly. Because of the economic and political need for slow transitions, which help reduce adjustment costs and political opposition, the efficiency trade-off from using an iterative approach is small. Importantly, the information demands required for an iterative approach (involving *average* local damage over several time periods) are much smaller than for a one-step approach (which requires quantifying *marginal* local damage).

The results in this paper are based on three specific functional forms for firms' marginal abatement cost, but these forms allow for an *extensive margin* of pollution abatement in addition to the usual *intensive margin* of pollution abatement. The proposed permit market remains feasible and optimal even when firms face all-or-nothing abatement decisions (e.g., install a scrubber or not) and marginal abatement cost is not defined. At the plant level, abatement cost functions are often non-differentiable step functions.

The novel hybrid market introduced in this paper solves several key challenges for tackling pollution 'hot spots' by simplifying the notion of trading ratios. It also extends to temporal clustering of emissions by suitably modifying the hazard function across different time periods (seasons). The market is informationally simple and transparent for regulators and firms alike. It is robust to different types of abatement behavior of firms. It only relies on information about average damage and does not require assumptions about marginal damage. Iteration over time allows marginal damage to be approximated by changes of average damage over time. By consolidating all transactions, the single permit market maximizes transaction liquidity, preserves anonymity, and reduces transaction costs. Lastly, the hybrid market is politically appealing by mandating a gradual adjustment process.

Appendix 1: Quasi-linear optimization

This paper makes use of a novel approach to solve the problem of optimizing a non-linear policy objective function. There is a class of policy problems that are characterized by (a) a policy objective function that is non-linear (e.g., a dose–response function); (b) a macro-economic policy instrument (e.g., a country-wide emission cap) and a set of corresponding micro-economic policy instruments (e.g., firm-level price factors); (c) additive separability of the micro-economic instruments; (d) monotonicity of the effect of the macro-economic variable with respect to the objective function; and (e) the ability to use the macro-economic policy instrument to drive a gradual adjustment process over time. The “quasi-linear optimization” approach introduced in this paper replaces a complex non-linear optimization problem (which yields an optimal solution in a single step) with an equivalent quasi-linear optimization problem that is much easier to solve, but only through gradual convergence to the optimum over time. Specifically, consider the problem of minimizing the single-valued quadratic objective function

$$F(\mathbf{x}, z) = \left[\sum_i f_i(x_i, z) \right]^2 + \sum_i g_i(x_i, z) \\ \equiv f(\mathbf{x}, z)^2 - g(\mathbf{x}, z)$$

where $\mathbf{x} = (x_1, \dots, x_n)$ is a vector of n micro-economic policy instruments that satisfy $\partial^2 F / \partial x_i \partial x_j = 0$, z is a macro-economic policy instrument that satisfies $\partial F / \partial z < 0$ for any $z > z^*$, and α is a scaling parameter. Many non-linear policy

problems can be approximated reasonably well as quadratic functions, at least near the optimal solution. Solving the system of first-order conditions

$$\forall i : f(\mathbf{x}, z) \frac{\partial f_i}{\partial x_i} + \frac{\partial g_i}{\partial x_i} = 0$$

and

$$f(\mathbf{x}, z) \sum_i \frac{\partial f_i}{\partial z} + \sum_i \frac{\partial g_i}{\partial z} = 0$$

yields the optimal policy (\mathbf{x}^*, z^*) . However, this system of $n + 1$ equations in $n + 1$ unknowns can be very difficult to solve. Even though the partial derivatives may only depend on x_i separately, the appearance of $f(\mathbf{x}, z)$ interacts the x_i in each equation with all other $x_{j \neq i}$.

Quasi-linear optimization involves minimizing the objective function

$$\tilde{F} \equiv f(\mathbf{x}_{t-1}, z_{t-1})f(\mathbf{x}_t, z_t) + g(\mathbf{x}_t, z_t)$$

gradually over time t . Objective function \tilde{F} approximates F when $|z_t - z_{t-1}|$ is small. In each period t , the n first-order conditions $\partial f_i / \partial x_{t,i} = 0$ deliver a set of micro-economic policies \mathbf{x}_t^* that are optimal in that time period conditional on the choice of macro-economic policy z_t that is governed exogenously. Because of the independence of the micro-economic policy instruments, these n first-order equations can be solved individually for the optimal policy $x_{i,t}^*$ at time t . Note that $f(\mathbf{x}_{t-1}, z_{t-1})$ is simply a scalar evaluated in the previous time period. The policy maker gradually reduces z_t over time, starting from the initial policy-less state z_0 . Decreasing z_t in small steps, the optimization procedure stops when decreasing z_t yields no further reduction in $\tilde{F}(\mathbf{x}_t^*, z_t)$, i.e., when $\partial \tilde{F}(\mathbf{x}^*, z) / \partial z = 0$.

Appendix 2: Pollution-plant centrality

The regulator’s main function in overseeing the permit market is to announce new $\phi_{i,t} = Q_{i,t} / \bar{Q}_{i,t}$ at the beginning of each trading period t , based on measurements from the previous period. As ambient concentrations A_j depend on contributions from all plants, the hazard summation factor $Q_{i,t}$ can be written as

$$\begin{aligned} Q_{i,t} &= \omega \sum_j P_j A_{j,t-1} d_{ij} \\ &= \omega^2 \sum_j P_j d_{ij} \sum_k d_{kj} E_{k,t-1} \end{aligned}$$

Emissions from other plants affect plant i through the network of dispersion coefficients d_{ij} . Firms are interconnected participants in the permit market where actions by one firm such as increasing or decreasing emissions may influence neighbouring plants by raising or lowering their particular hazard factor ϕ_i .

It is possible to capture the degree of interconnectedness through empirical measures of centrality. Such measures are widely used in the theory of social and economic networks (Jackson 2008), and they are also used in applications such as Google's web page ranking. In matrix notation, $\mathbf{Q} = \omega^2 \mathbf{D} \mathbf{P} \mathbf{D}^T \mathbf{E}$, where \mathbf{Q} and \mathbf{E} are $n \times 1$ vectors of each plant's (non-normalized) hazard factor and emission level, \mathbf{D} is an $n \times m$ matrix of dispersion coefficients, \mathbf{P} is an $m \times m$ diagonal matrix with region factors P_j , and the T superscript indicates the matrix transpose. $\mathbf{S} \equiv \mathbf{D} \mathbf{P} \mathbf{D}^T$ is a symmetric $n \times n$ matrix that captures the inter-firm influences weighted by \mathbf{P} . Matrix \mathbf{S} can be normalized through the transformation $\mathbf{U} \equiv \mathbf{R} \mathbf{S} \mathbf{R} - \mathbf{I}$ where $\mathbf{R} = [\text{diag}(\mathbf{S})]^{-1/2}$ and \mathbf{I} is the identity matrix; this transformation is the same procedure that is used to compute a Pearson correlation matrix from a variance-covariance matrix. Thus, all elements in matrix \mathbf{U} identify the relative impact of one plant on another, scaled to the range $[0, 1]$. Subtracting \mathbf{I} ensures that the diagonal elements in \mathbf{U} are all zero, as is required for adjacency matrices that are used to analyze networks.

The influence of an individual plant in the emission market network can be captured by eigenvector centrality, originally proposed by Bonacich (1972). The largest eigenvalue $\bar{\lambda}$ that solves the eigenvector equation $\mathbf{U} \mathbf{v} = \lambda \mathbf{v}$ is associated with the principal eigenvector $\bar{\mathbf{v}}$, which can be obtained numerically through power iteration even for very large matrices. The centrality measure can be suitably disaggregated into plants with zero or positive centrality. For the $n^+ \leq n$ firms with positive centrality, rescaling makes the measures easier to interpret. Let $\chi_i \equiv (\bar{v}_i / \sum_k \bar{v}_k) / (1/n^+)$ denote plant centrality with an average of 1. Values larger than 1 indicate above-average centrality, and values below 1 indicate below-average centrality.

The plant-specific hazard factors ϕ_i will be strongly rank-correlated with the centrality measures χ_i . The firm centrality measure is useful in the context of spatially interacting plants because it readily identifies which plants affect their neighbours more, and which less. For example, a centrality measure of $\chi_i = 1.5$ suggests that plant i has 50 % stronger influence on neighbouring plants than the average plant. By comparison, a plant with $\chi_i = 0.5$ is much less spatially connected, influencing other firms 50 % less than average. Plants that are far from any other plants have a χ_i equal or close to zero.

References

- Akao K, Managi S (2013) A tradable permit system in an intertemporal economy: a general equilibrium approach. *Environ Resour Econ* 55(3):309–336. doi:[10.1007/s10640-012-9628-5](https://doi.org/10.1007/s10640-012-9628-5)
- Antweiler W (2003) How effective is green regulatory threat? *Am Econ Rev* 93(2):436–441. doi:[10.1257/000282803321947489](https://doi.org/10.1257/000282803321947489)
- Atkinson SE, Tietenberg TH (1982) The empirical properties of two classes of designs for transferable discharge permits. *J Environ Econ Manag* 9(2):101–121. doi:[10.1016/0095-0696\(82\)90016-X](https://doi.org/10.1016/0095-0696(82)90016-X)
- Baumol WJ, Oates WE (1988) *The theory of environmental policy*, 2nd edn. Cambridge University Press, Cambridge
- Benbear LS, Stavins RN (2007) Second-best theory and the use of multiple policy instruments. *Environ Resour Econ* 37(1):111–129. doi:[10.1007/s10640-007-9110-y](https://doi.org/10.1007/s10640-007-9110-y)

- Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. *J Math Sociol* 3:113–120. doi:[10.1080/0022250X.1972.9989806](https://doi.org/10.1080/0022250X.1972.9989806)
- Burtraw D (2013) The institutional blind spot in environmental economics. *Dædalus* 142(1):110–118. doi:[10.1162/DAED_a_00188](https://doi.org/10.1162/DAED_a_00188)
- Cronshaw MB, Kruse JB (1996) Regulated firms in pollution permit markets with banking. *J Regul Econ* 9(2):179–189
- Egenhofer C (2003) The compatibility of the kyoto mechanisms with traditional environmental instruments. In: Carraro C, Egenhofer C (eds) *Firms, governments and climate policy: incentive-based policies for long-term climate change*. Edward Elgar, Cheltenham, pp 17–82 chap 1
- Farrow RS, Schultz MT, Celikkol P, Van Houtven GL (2005) Pollution trading in water quality limited areas: use of benefits assessment and cost-effective trading ratios. *Land Econ* 81(2):191–205. doi:[10.3368/le.81.2.191](https://doi.org/10.3368/le.81.2.191)
- Fowlie M, Muller N (2013) Market-based emissions regulation when damages vary across sources: what are the gains from differentiation?. NBER Working Paper 18801
- Fowlie M, Reguant M, Ryan SP (2010) Pollution permits and the evolution of market structure, working Paper, Massachusetts Institute of Technology
- Fowlie M, Reguant M, Ryan SP (2015) Market-based emissions regulation and industry dynamics. *J Polit Econ* 123 (**forthcoming**)
- Førsund FR, Nævdal E (1998) Efficiency gains under exchange-rate emission trading. *Environ Resour Econ* 12(4):403–423. doi:[10.1023/A:1008247511950](https://doi.org/10.1023/A:1008247511950)
- Gangadharan L (2000) Transaction costs in pollution markets: an empirical study. *Land Econ* 76(4):601–614. doi:[10.2307/3146955](https://doi.org/10.2307/3146955)
- Goodkind A, Coggins JS, Marshal JD (2014) A spatial model of air pollution: the impact of the concentration-response function. *J Assoc Environ Resour Econ* 1:451–479. doi:[10.1086/678985](https://doi.org/10.1086/678985)
- Hoel DG, Portier CJ (1994) Nonlinearity of dose–response functions for carcinogenicity. *Environ Health Perspect* 102(Suppl 1):109–113
- Holland SP, Yates AJ (2015) Optimal trading ratios for pollution permit markets. *J Public Econ* 125:16–27. doi:[10.1016/j.jpubeco.2015.03.005](https://doi.org/10.1016/j.jpubeco.2015.03.005)
- Hung MF, Shaw D (2005) A trading-ratio system for trading water pollution discharge permits. *J Environ Econ Manag* 49(1):83–102. doi:[10.1016/j.jeem.2004.03.005](https://doi.org/10.1016/j.jeem.2004.03.005)
- Jackson MO (2008) *Social and economic networks*. Princeton University Press, Princeton, NJ
- Konishi Y, Coggins JS, Wang B (2015) Water quality trading: can we get the prices of pollution right? *Water Resour Res* 51(5):3126–3144. doi:[10.1002/2014WR015560](https://doi.org/10.1002/2014WR015560)
- Krysiak FC, Schweitzer P (2010) The optimal size of a permit market. *J Environ Econ Manag* 60(2):133–143. doi:[10.1016/j.jeem.2010.05.001](https://doi.org/10.1016/j.jeem.2010.05.001)
- Lee TC, Chen HC, Liu SM (2013) Optimal strategic regulations in international emissions trading under imperfect competition. *Environ Econ Policy Stud* 15(1):39–57. doi:[10.1007/s10018-012-0033-7](https://doi.org/10.1007/s10018-012-0033-7)
- Montero JP (1997) Marketable pollution permits with uncertainty and transaction costs. *Resour Energy Econ* 20(1):27–50. doi:[10.1016/S0928-7655\(97\)00010-9](https://doi.org/10.1016/S0928-7655(97)00010-9)
- Montgomery DW (1972) Markets in licenses and efficient pollution control programs. *J Econ Theory* 5(3):395–418. doi:[10.1016/0022-0531\(72\)90049-X](https://doi.org/10.1016/0022-0531(72)90049-X)
- Muller NZ, Mendelsohn R (2009) Efficient pollution regulation: getting the prices right. *Am Econ Rev* 99(5):1714–1739. doi:[10.1257/aer.99.5.1714](https://doi.org/10.1257/aer.99.5.1714)
- Newell R, Pizer W, Zhang J (2005) Managing permit markets to stabilize prices. *Environ Resour Econ* 31(2):133–157. doi:[10.1007/s10640-005-1761-y](https://doi.org/10.1007/s10640-005-1761-y)
- Rindi B (2008) Informed traders as liquidity providers: anonymity, liquidity, and price formation. *Rev Finance* 12(3):497–532. doi:[10.1093/rof/rfm023](https://doi.org/10.1093/rof/rfm023)
- Roberts MJ, Spence M (1976) Effluent charges and licenses under uncertainty. *J Public Econ* 5(3–4):193–208
- Rubin JD (1996) A model of intertemporal emission trading, banking, and borrowing. *J Environ Econ Manag* 31(3):269–286. doi:[10.1006/jjeem.1996.0044](https://doi.org/10.1006/jjeem.1996.0044)
- Schennach SM (2000) The economics of pollution permit banking in the context of title IV of the 1990 Clean Air Act Amendments. *J Environ Econ Manag* 40(3):189–210. doi:[10.1016/0022-0531\(72\)90049-X](https://doi.org/10.1016/0022-0531(72)90049-X)
- Schmalensee R, Stavins RN (2012) The SO₂ allowance trading system: The ironic history of a grand policy experiment. Tech. rep., MIT Center for Energy and Environmental Policy Research, CEEPR WP 2012-012

- Schmalensee R, Stavins RN (2013) The SO₂ allowance trading system: the ironic history of a grand policy experiment. *J Econ Perspect* 27(1):103–122. doi:[10.1257/jep.27.1.103](https://doi.org/10.1257/jep.27.1.103)
- Siikamäki J, Burtraw D, Maher J, Munninge C (2012) The U.S. Environmental Protection Agency's Acid Rain Program. Tech. rep., Resources for the Future
- Theissen E (2003) Trader anonymity, price formation, and liquidity. *Rev Finance* 7(1):1–26. doi:[10.1023/A:1022579423978](https://doi.org/10.1023/A:1022579423978)
- Tietenberg TH (2006) Emissions trading: principles and practice, 2nd edn. RFF Press, Washington, DC
- Tietenberg TH, Grubb M, Michaelowa A, Swift B, Zhang Z (1999) International rules for greenhouse gas emissions trading: defining the principles, modalities, rules and guidelines for verification, reporting and accountability. UNCTAD, Geneva
- US-EPA (2004) EPA's risk-screening environmental indicators (RSEI) chronic human health methodology, version 2.1.2, Appendix A. Tech. rep., United States Environmental Protection Agency. URL <http://www.epa.gov/opptintr/rsei/pubs/>
- Welker M (1995) Disclosure policy, information asymmetry, and liquidity in equity markets. *Contemp Account Res* 11(2):801–827. doi:[10.1111/j.1911-3846.1995.tb00467.x](https://doi.org/10.1111/j.1911-3846.1995.tb00467.x)
- Yap D, Reid N, de Brou G, Bloxam R (2005) Transboundary air pollution in Ontario. Tech. rep., Ontario Ministry of the Environment