Managing Pollution Risk through Emissions Trading

Gaurav Ghosh and James Shortle

January 2012
Managing Pollution Risk through Emissions Trading

January 2012

Authors’ addresses:

Gaurav Ghosh
Institute for Future Energy Consumer Needs and Behavior (FCN)
School of Business and Economics / E.ON Energy Research Center
RWTH Aachen University
Mathieustrasse 6
52074 Aachen, Germany
E-mail: gghosh@eonerc.rwth-aachen.de

James Shortle
Director, Environmental and Natural Resources Institute
Department of Agricultural Economics and Rural Sociology
Pennsylvania State University
112C Armsby,
University Park, PA 16802, U.S.A.
E-mail: jshortle@psu.edu

Publisher:  Prof. Dr. Reinhard Madlener
Chair of Energy Economics and Management
Director, Institute for Future Energy Consumer Needs and Behavior (FCN)
E.ON Energy Research Center (E.ON ERC)
RWTH Aachen University
Mathieustrasse 6, 52074 Aachen, Germany
Phone: +49 (0) 241-80 49820
Fax: +49 (0) 241-80 49829
Web: www.eonerc.rwth-aachen.de/fcn
E-mail: post_fcn@eonerc.rwth-aachen.de
Managing Pollution Risk through Emissions Trading

Gaurav Ghosh\textsuperscript{a}, James Shortle\textsuperscript{b}

\textsuperscript{a} Institute for Future Energy Consumer Needs and Behavior (FCN), School of Business and Economics / E.ON Energy Research Center, RWTH Aachen University, Mathieustrasse, 52074 Aachen, Germany

\textsuperscript{b} Department of Agricultural Economics and Rural Sociology, Pennsylvania State University, 112C Armsby Building, University Park, PA 16802, USA

January 2012

Abstract

We compare two tradable permit markets in their ability to meet a safety first environmental target at least cost when some polluters have stochastic, correlated, and non-measurable emissions. In both markets, the point source permit defines the allowable level of the observed (deterministic) point source pollution load. The permit for unobservable and stochastic nonpoint source pollution cannot be defined in this way. One market bases the nonpoint permit on expected nonpoint pollution and uses a trading ratio between the two pollution types to manage stochasticity. This model follows existing point-nonpoint markets for water quality trading. The second model defines the nonpoint permit as a multi-attribute good, where the attributes inform the market about the stochasticity of the underlying pollution load. The multi-attribute permit market is demonstrated to out-perform the trading ratio market. This result is an artifact of polluters directly pricing stochasticity in the former market but not in the latter, where stochasticity is only controllable under highly restrictive conditions.

JEL Classification Codes: D02, D62, D81, Q52, Q53, Q58

Keywords: Emissions trading, Environmental policy; Market design; Nonpoint pollution; Water Quality

1 Introduction

Beginning with the seminal works of Crocker (1966), Dales (1968) and Montgomery (1972), a large literature has developed on the use of emissions trading to achieve environmental

\textsuperscript{*} Corresponding author. Tel. +49 241 80 49 820, fax. +49 241 80 49 829, RMadlener@eonerc.rwth-aachen.de (R. Madlener)
targets. The economic case for emissions trading is that it can achieve environmental targets at lower social cost than emissions taxes and traditional design and performance standards (Tietenberg, 1990, 2003). Real world success stories for air emissions trading, such as the US Environmental Protection Agency’s (US EPA’s) Acid Rain Control Program (Joskow et. al., 1998), have spurred interest in expanding the scope of emissions trading to other air pollutants (e.g., carbon) and other media (e.g., water) (Organization for Economic Cooperation and Development (OECD), 2004).

Water quality trading was the focus of Dales (1968), which recommended the market as a tool for environmental management. However, it has not until recently been a focus of applications of the tool. Interest is now high. Dozens of water quality trading initiatives are underway for nutrients, sediments and other pollutants in the US and other countries (Breetz et al., 2004; Selman et al., 2009). Nonpoint sources (NPS) are significant sources of pollutants targeted by these emerging trading programs. Yet, nonpoint water pollution is in fundamental ways an unlikely candidate for application of this mechanism. Idealized “textbook” models of emissions trading require that emissions (1) can be accurately metered for each regulated emitter, (2) are substantially under control of the polluter, and (3) that the spatial location of emissions within the market does not affect environmental outcomes (e.g., Ellerman, 2005). The textbook model is best approximated in practice by the cap-and-trade market for sulfur dioxide and nitrogen oxide emissions in the US (e.g., Schmalensee et al., 1998), and the European Union Emissions Trading System (EU ETS).

The first two requirements of the textbook model are necessary to meaningfully define tradable property rights in actual emissions. Neither of these requirements is satisfied by nonpoint pollution. The diffuse and complex pathways that pollutants follow from NPS to water resource make routine and accurate metering of nonpoint pollution prohibitively expensive, thus
violating the first assumption. Further, weather plays a large role in determining the volume, timing, and form of water pollutants from nonpoint sources. This introduces a significant random component, and implies that timings and levels of discharges are not completely under control of nonpoint managers, thus violating the second assumption. Efforts to reduce nonpoint pollution are appropriately viewed as improving the probability distribution of the pollution load rather than its actual level (Horan and Ribaudo, 1999; Segerson, 1988; Shortle and Dunn, 1986).

These features of nonpoint pollution raise two significant and related issues for market design. The first is what to define as the NPS permit or credit. The second is how to design trading rules to manage the probability that imperfect monitoring and control result in an adverse ex post environmental outcome. The uncertainty surrounding this probability is exacerbated when NPS emissions are correlated, as is often the case (Merz and Blöschl, 2009; Kampas and White, 2003). The current approach to water quality trading in programs is to define the tradable nonpoint permit in terms of estimated abatements based on hydrological water quality models (see Breetz et al. (2004) for examples). However, because the prediction errors in water quality models are large (Stow et al., 2007; Gitau and Veith, 2006; National Research Council (NRC), 2001; Reckhow, 1994, 2003), there is enormous uncertainty about the actual water quality outcomes of individual trades based on modeled emissions.

This uncertainty is addressed by the application of an “uncertainty” trading ratio to trades with nonpoint polluters, which essentially discounts the effectiveness of NPS pollution abatements. The ratio is typically defined in terms of the units of abatement in nonpoint pollution required to offset a unit increase of point source (PS) pollution. Uncertainty ratios in US water quality trading programs range from 1.1:1 to 4:1 (Morgan and Wolverton, 2005). This paper will demonstrate that the use of modeled mean pollution loads as an approximation of nonpoint pollution and the use of trading ratios to address the risk of adverse outcomes is suboptimal, and
will offer an alternative market design that offers greater promise for cost-efficiency in water quality management through trading.

We develop the concept of a safety first pollution constraint in the context of the management of stochastic pollution and derive necessary conditions for compliance with this constraint at least cost. The safety first constraint acknowledges that pollution cannot be controlled deterministically, and thus that violations of a limit on pollution may occur with a non-trivial frequency due to stochastic processes beyond the control of polluters. The safety first decision criterion has three characteristics that make it suitable for stochastic pollution control. First, it is consistent with US water quality regulations that recognize the uncertainty in water quality outcomes, such as the total maximum daily load (TMDL) requirement. Second, it is intuitively comprehensible as “equivalent to the use of significance levels for statistical decision making” and finally, it is practical and complements a wide range of risk assessment models (Lichtenberg and Zilberman, 1988). For these reasons, safety first rules will be widely understood and accepted by policy makers with a wide range of technical and economic backgrounds.

We propose an innovative market design and show that the competitive equilibrium under this design satisfies the safety first constraint at least cost. The primary innovation in market rules, when compared to conventional tradable permit markets like the EU ETS or US water quality markets with modeled-mean emissions-based permits, is in the definition of the nonpoint permit. Here, the nonpoint permit has two attributes: the mean and variance of the underlying abatement in emissions. Through this definition, market participants are fed information on the underlying uncertainty associated with the abatement. Other market rules, in the form of

---

1 The TMDL requires emissions or pollution load reductions to achieve water quality goals in impaired waters and includes a margin of safety to account for seasonal variations and other sources of uncertainty (Tetra Tech Inc., 1999; Ohio Environmental Protection Agency, 2002).
probabilistic pollution caps based on safety first design, incentivize participants to control this uncertainty. We also model trading ratio markets and show that, in general, these markets will not meet the safety first constraint at least cost. Reasons include an excessive information burden on the regulatory authority and misaligned incentives for the polluters.

The next section develops the safety first pollution constraint and obtains conditions on polluter behavior required for its least cost attainment. While designing the constraint, we consider the correlations between stochastic emissions and their control. We then develop the market with the multi-attribute nonpoint abatement (MANA) permits and derive necessary conditions for the competitive market equilibrium. We examine restrictions that ensure that this equilibrium is safety first compliant at least cost. A trading ratio market, which is conceptually identical to water quality trading markets in current use, is also analogously developed and analyzed. Finally, the two markets are compared and the results put in a broader policy context. Throughout, we draw parallels with existing regulation for water quality markets in Pennsylvania to illustrate how the MANA market might work in the real world.

2 Optimality under Safety First Pollution Policy Design

Although our model is broadly applicable to all stochastic pollutants, we frame our discussion in terms of water quality trading because it is in this domain that much of the research and policy discussion on stochastic nonpoint pollution takes place. Following the US EPA’s TMDL approach (see USEPA, 2003) we take the environmental objective to be the limiting of the aggregate pollution from point and nonpoint sources in a watershed to a level consistent with the achievement of an in-stream water quality goal, with some margin of safety. We formally define this objective as a safety first pollution constraint.
The safety first constraint is a probabilistic statement expressed in terms of the first and second order moments of the random factors to be controlled (e.g. Beavis and Walker, 1983; Lichtenberg and Zilberman, 1988). When the factors are independent these moments consist of their means and variances. When correlated, covariances are also included. The latter is the more complicated scenario, given the greater number of control variables. The science indicates that it is also the relevant assumption for our problem. Contiguous polluters in a watershed tend to have correlated runoffs because of similar precipitation, soil moisture, stream network density and other elements (Merz and Blöschl, 2009; Skoien and Blöschl, 2007).

However, in the context of point-nonpoint trading even though emissions from nonpoint sources are correlated, these correlations need not be explicitly controlled, i.e. independence may be assumed. This follows from how nonpoint pollution control technologies; referred to in water quality law, and in technical and policy literature on nonpoint pollution control as Best Management Practices (BMPs); curtail pollution. To illustrate, consider a set of farms contributing to nitrogen pollution of a stream, lake, or estuary. Their correlated emissions may be represented as a multivariate probability distribution. The emissions of a particular farm will then be represented by the corresponding marginal distribution. The marginal distribution will depend on the technology implemented. For example, a riparian buffer can reduce nitrogen loads into agricultural watersheds by more than 80 percent (Hefting and de Klein, 1998; Osborne and Kovacic, 1993), while the effectiveness of a nutrient management plan is in the 30-40 percent range (Hall and Risser, 1993; van Dyke et al., 1999). Even during extreme storm events, a well-designed buffer or fertilizer applicator retains much of its effectiveness and reduces pollution loads significantly (Lee et al., 2003; Ressler et al., 1998). These BMPs dampen the

---

2 We focus on agriculture as an illustration because it is the most significant source of NPS pollution in the US (NRC, 1992; USEPA, 1996), and the most important potential participant in existing and planned point-nonpoint trading programs (Ribaudo and Gottlieb, 2011).
farm’s emissions response to stochastic events, by making extreme surges infeasible. Abatement technologies or BMPs, in effect, truncate the emissions distribution.

Truncation affects the variance of emissions and its correlations (or equivalently, covariances) with other emissions in two ways. First, it attenuates the variance and all correlations. Beyond a certain level of technological efficiency, correlations attenuate to such a degree that the pollution loads become approximately independent. Second, covariances attenuate at a faster rate than the variance, i.e. covariances associated with a given nonpoint’s emissions approach zero at lower levels of truncation in the joint emissions distribution than the variance. Aitken (1964) establishes these results for the specific case of a truncated bivariate normal distribution. We illustrate them for the more general multivariate normal distribution through a simulation in the Appendix where of a set of correlated polluters implementing BMPs of different efficiencies are modeled.3

Given the relationships between truncation, variance and correlation, it follows that the effectiveness of a BMP in reducing the stochasticity of individual and aggregate NPS emissions can be gauged by only focusing on how it affects the variance. The Environmental Agency (EA) does not need to explicitly control correlations among nonpoint polluters since the BMP will, in any case, reduce correlation between nonpoint emissions faster than it will reduce the variance. It can assume that nonpoint pollution loads are independent for the purpose of policy design.

This is convenient from the practitioner’s point of view for two reasons. First, existing studies on the efficiency of BMPs only focus on how they reduce onsite emissions. There is no

---

3 There are conflicting views on the appropriate distribution for non-negative environmental variables (like emissions). In low-information or high uncertainty scenarios, a nonparametric approximation of the data is suitable (Wets, 1983). In other scenarios, the multivariate normal distribution is used, and is justified through the Central Limit Theorem. Or, skewed probability distributions with a positive support are used, of which the lognormal is most common (Parkin and Robinson, 1992). Our results on the truncated multivariate normal distribution hold for the multivariate lognormal and other distributions like multivariate gamma and multivariate skew-normal since correlations under these distributions are monotonic transformations of correlations under the multivariate normal distribution (Rendu, 1979; Azzalini and Dalla Valle, 1996; Krishnaiah and Rao, 1961).
literature on how BMPs affect correlation between onsite pollution loads and loads at neighboring locations. Hence, while it is feasible to use time series of emissions before and after the implementation of the BMP to estimate variance, reliable estimates of correlation are unavailable given the current state of the science. Second, there is an incentive problem that may be insurmountable. Controlling correlation requires cooperation between the two affected parties. An NPS will have to coordinate its abatement strategy with all other correlated polluters. Apart from hidden action and hidden information complications, the complexity of the decision problem will impede market efficiency.\textsuperscript{4}

Let \( L \) be aggregate pollution and let \( \bar{L} \) be an exogenously chosen watershed-level pollution target. In current programs \( \bar{L} \) is defined by the TMDL or an analogous regulatory driver drawn from state or federal legislation on water quality mandates or emissions discharge limits. \( \bar{L} \) may be defined in terms of upper bounds on monthly or annual discharges, or concentrations of specified pollutants (see Breetz et al. (2004) for examples). Following the safety first literature, the pollution constraint is expressed probabilistically. The EA is tasked with designing and enforcing the constraint that the probability that aggregate pollution exceeds the target be less than \( \alpha \), i.e. \( \Pr(L \geq \bar{L}) \leq \alpha \) where \( \alpha \) is an exogenously fixed probability. It is the highest probability of policy failure deemed acceptable by a society or community in an uncertain environment. It may be interpreted as an index of societal tolerance: a high \( \alpha \) would imply that the society has a high tolerance for policy failure. The watershed has \( I \) nonpoint sources and \( J \) point sources, indexed by \( i \) and \( j \) respectively. For convenience, \( I \) and \( J \) will also refer to the sets of nonpoint and point sources. Let \( r_i \in \mathbb{R}_+ \) be NPS \( i \)'s stochastic pollution load or emissions and

\textsuperscript{4} Although we justify and assume that NPS emissions are independent, the assumption is not critical to our results. The results are identical to those in a related paper where correlations were explicitly controlled, but where technological and strategic impediments were not considered (see Ghosh and Shortle, 2010).
let $e_j \in \mathbb{R}_+$ be PS $j$’s deterministic load into the watershed in a trading period.\(^5\) Let $\mu_i \in \mathbb{R}_+$ be the mean and let $\sigma_i^2 \in \mathbb{R}_{++}$ be the variance of $r_i$. Together, $\mu_i$ and $\sigma_i^2$ supply a second order description of the probability distribution of $i$’s load. These moments may be estimated from measurements and statistics calculated from soil transport and hydrological models that map land use practices influencing nonpoint emission flows (e.g. Sprague, 2001; Stow et al., 2001).

True (but unknown) aggregate pollution in the watershed is $L = \sum_i r_i + \sum_j e_j$. The EA’s safety first pollution constraint is $\Pr(\sum_i r_i + \sum_j e_j \geq \bar{L}) \leq \alpha$. If the joint distribution of emissions is known and its “deterministic equivalent” exists, then the probability statement above can be expressed exactly in terms of means, variances, covariances and other moments of the emissions distribution (Beavis and Walker, 1983; Wets, 1983; Kampas and White, 2003). But given the non-measurability of nonpoint emissions and the prediction errors in current water quality models, this distribution is unknown and the conditions for a deterministic equivalent are not satisfied. A common nonparametric approximation of probability statements when the distribution of the random variable is unknown (or a deterministic equivalent does not exist) is based on Chebychev’s inequality (Wets, 1983). By this inequality the statement $\Pr(L \geq E(L) + \sqrt{V(L)/\alpha}) \leq \alpha$ is always true, where $E(L) = \sum_i \mu_i + \sum_j e_j$ and $V(L) = \sum_i \sigma_i^2$ are the mean and variance of $L$. Using the Chebychev inequality, the probabilistic statement that is the safety first constraint can be expressed as

\(\footnote{We interpret PS and NPS emissions as the amount delivered from the location of the PS or NPS to a receptor rather than as emissions onsite. This interpretation is consistent with emerging trading programs in the US. To illustrate, Pennsylvania’s trading program uses fixed edge-of-segment ($EOS$) and delivery ($\delta$) ratios to calculate the contributions of individual polluters to specific receptors downstream (PADEP, 2007). The contributions are then used to define permits. A watershed is divided into subwatersheds, each associated with one or more $EOS$ and $\delta$ ratios that are differentiated by crop or production factors. Consider PS $j$ in subwatershed $x$. If $j$’s on-site emissions are $\bar{\delta}_j$ then emissions at the receptor point are $e_j = \bar{\delta}_j \times EOS_x \times \delta_x$. Note that under these regulations, emissions transport is assumed to be deterministic. In truth, transport is stochastic because of limited information about its process. We assume deterministic transport to allow greater comparability between our designed model and existing markets. We aim to show how minor tweaks to the existing paradigm, as exemplified by the Pennsylvanian program, can cause better and cost-effective environmental outcomes.}
\[
\sum_i \mu_i + \sum_j e_j + \sqrt{\frac{\sum_i \sigma_i^2}{\alpha}} \leq \bar{L}
\]

We omit discussion of the derivation of safety first constraint (1) since it draws on earlier literature (e.g. Beavis and Walker, 1983) except to note that \( \sqrt{V(L)/\alpha} \) may be interpreted as an uncertainty penalty imposed by the EA and inversely proportional to societal tolerance for policy failure. As concern about pollution increases, the tolerance for failure in adhering to the pollution target \( \bar{L} \) falls. An EA representing societal mandates accommodates this decreased tolerance by decreasing \( \alpha \), which inflates the uncertainty penalty \( \sqrt{V/\alpha} \), and its marginal effect \( 1/2\sqrt{\alpha V} \). To counteract this inflation, which is necessary for the non-violation of (1), polluters are forced to reduce elements in \( \{\mu_i, \sigma_i^2\}_{i \in I} \) or \( \{e_j\}_{j \in J} \). The relative effects of these reductions are mediated by \( \alpha \) and \( V \). Marginal reduction of \( \mu_i \) or \( e_j \) weaken (1) by one unit while a marginal reduction in \( \sigma_i^2 \) weakens it by \( 1/2\sqrt{\alpha V} \) units. It follows that control of \( \sigma_i^2 \) is less important than control of \( \mu_i \) or \( e_j \) if \( V > 1/4\alpha \) and vice versa. This inequality is more likely to hold when societal tolerance is high (i.e. \( \alpha \) is low). This makes intuitive sense: if tolerance for ex post policy violation is high then there is less incentive to control \( \sigma_i^2 \) for any \( i \in I \).

Similar to the “textbook” trading model we assume that point sources can control their measured emissions, \( e_j \), and nonpoints can control their mean emissions, \( \mu_i \). Extending the model, we also assume that nonpoints can adjust the emissions variance \( \sigma_i^2 \). Furthermore, we assume that in a given range, they can adjust these variables flexibly. Commonly used in safety first models, this assumption is plausible in a scenario where an NPS has access to a large set of BMPs and implements a subset of them. We can think of each technology as having a feasible region in \( \{\mu_i, \sigma_i^2\} \) space. When multiple technologies are implemented, this region expands. By
flexibility in choice of $\mu_i$ and $\sigma_i^2$, we merely imply that by changing technological parameters and subject to stochastic effects, the nonpoint source can reach all points in this feasible region.

Consider a farmer participating in Pennsylvania’s water quality trading (WQT) market, the largest point-nonpoint trading program implemented to date, measured by the number of potential participants and the size of the regulated area (OECD, 2012). Not only can the farmer choose from about 20 abatement technologies, but each technology can be implemented in many different ways (PADEP, 2007, 2008). For example, when implementing a riparian buffer, he must first decide on its vegetative composition. He could choose a grass, shrub or tree buffer, or some combination thereof. For each type, he must choose a combination of species. The choice of species will be affected by climate, longevity, nutrient uptake, maintenance requirements and other factors. Next, numerous planting decisions will have to be made, including timing, buffer width and costs. Analogously, implementing other abatement technologies are complex decisions. Given this complex decision environment, where multiple technologies may be implemented and where changing some parameters (like buffer width) have marginal effects, we assume that the farmer can control his loadings such that $\mu_i$ and $\sigma_i^2$ vary independently.

Let $\pi_i(\mu_i, \sigma_i^2)$ and $\pi_j(e_j)$ be the restricted production profit functions$^6$ for $i$ and $j$ respectively and that $\pi_i(0, \sigma_i^2) = \pi_i(\mu_i, 0) = \pi_j(0) = 0$. The latter is a positive pollution condition, implying that point and nonpoint sources cannot produce without polluting. We assume that profits are increasing and concave in all arguments, i.e. (suppressing indices and arguments) $\pi' \geq 0$ and $\pi'' < 0$. This is equivalent to assuming that abatement – through reduction in $\mu_i$ or $\sigma_i^2$ for nonpoint sources and reduction in $e_j$ for point sources – is costly, and

---

$^6$ Restricted profit functions define the maximum profit associated with any emissions level, given that the input mix and output levels are chosen optimally. These functions are derived from a standard profit function, where profits are a function of inputs, production technology and prices. We do not supply a formal derivation, but refer interested readers to Graff Zivin and Small (2003). Restricted profit functions are useful because they allow modeling of profit as a function of the variables of interest.
that marginal cost increases as abatement, which may be conceptualized as negative emissions, increases. In the context of our Pennsylvanian market, this implies that polluters begin by investing in the cheaper BMPs. Once gains from these BMPs are exhausted they move onto more expensive BMPs. The joint profit for all polluters in the watershed is

\[
\sum_i \pi_i(\mu_i, \sigma_i^2) + \sum_j \pi_j(e_j)
\]  

(2)

Polluters jointly maximize profits by choosing \(\mu_i, \sigma_i^2\) and \(e_j\) such that (2) is maximized subject to (1) being satisfied. The constrained joint profit maximizing outcome also defines the lowest social cost of meeting the safety first constraint. The necessary conditions are

\[
\frac{\partial \pi_j}{\partial e_j} = \frac{\partial \pi_i}{\partial \mu_i} = 2\sqrt{\alpha V} \frac{\partial \pi_i}{\partial \sigma_i^2} = \lambda \quad \forall \quad i \in I, j \in J
\]  

(3)

By concavity and the positive profit condition, (3) is also sufficient for the least cost safety first allocation. This allocation has the following characteristics: First, marginal profits with respect to \(\mu_i\) and \(e_j\), \(\partial \pi_i / \partial \mu_i\) and \(\partial \pi_j / \partial e_j\), are equal to the shadow price of the safety first constraint, \(\lambda\), which is strictly positive by the Kuhn-Tucker conditions. This implies that \(\mu_i\) and \(e_j\) are marginally substitutable on a one-to-one basis and hence homogenous at the optimal allocation. Second, \(i^*\)’s marginal profits with respect to \(\sigma_i^2\), \(\partial \pi_i / \partial \sigma_i^2\), equals \(\lambda\) inflated by the marginal uncertainty penalty \(1/2\sqrt{\alpha V}\). These marginal profits may be interpreted as the EA’s valuation of \(\sigma_i^2\). Since the marginal uncertainty penalty is invariant across \(i\), \(\partial \pi_i / \partial \sigma_i^2\) is the same for all \(i\). The EA’s valuation of \(\sigma_i^2\) decreases as societal tolerance \(\alpha\) or total emissions variance \(V\) increases. Intuitively, individual variances are relatively less important factors in the safety first constraint as aggregate variance rises, and this reduces their value as control variables. Also, a community that tolerates policy failure due to uncertainty is less interested in
controlling said uncertainty through $\sigma_i^2$. Finally, $\sigma_i^2$’s value in terms of $\mu_i$ is $1/2\sqrt{\alpha V}$, which implies that the relative value of $\mu_i$ (or $e_j$) with respect to $\sigma_i^2$ increases as $\alpha$ or $V$ increase.

3 The MANA Market

We assume the market to be perfectly competitive. Point sources buy and sell permits to other point sources. In keeping with current regulations for point-nonpoint trading in the US, like the US EPA’s 2003 water quality trading policy, point sources buy permits from nonpoint sources and are held liable for the corresponding abatements (USEPA 2003; Selman et al., 2009). Nonpoint sources only sell permits. Although major contributors to water quality impairment, nonpoint sources (with the exception of large confined animal operations), unlike point sources, are not required to hold National Pollution Discharge Elimination System (NPDES) permits (USEPA, 2002). Without a legal requirement to control emissions they have no incentive to buy permits.

Let an allowance be a right to contribute to total emissions in the watershed, as measured at the receptor. It is a regulatory limit on emissions and is distinct from ex ante emissions levels. As is current practice (see Breetz et al., 2004 for examples), we assume that allowances are allocated by the EA or an empowered legislative or judicial authority, and known to the polluter at the start of the trading period. Let $\hat{\epsilon}_i$ and $\hat{\epsilon}_j$ be the allowances given to $i$ and $j$. They sum up to the watershed level emissions target, i.e. $\sum_j \hat{\epsilon}_j + \sum_i \hat{\epsilon}_i = \bar{L}$. The allowance is tradable and hence the MANA market is of the cap-and-trade type. The difference between PS $j$’s allowance and emissions is $a_j = \hat{\epsilon}_j - e_j$, referred to from here on as abatement.\(^7\) Since $e_j$ is accurately

\(^7\)Strictly speaking, one can differentiate between abatement when defined as the “difference between allowance and emissions” and “actual” abatement. The latter refers to the true level of emissions reduction. The former refers to some level of emissions reduction upon which permit levels are defined. Consider some $j$ with emissions $e_j^0$ prior to market entry. It is allotted some allowance $\hat{\epsilon}_j$. After entering the market it reduces emissions to some $e_j$. True
measurable, $a_j$ is known with certainty. All abatements are tradable and hence $j$’s permit supply is also $a_j$. NPS $i$’s abatement is the difference between its allowance and pollution load, $a_i = \hat{r}_i - r_i$. This abatement is unknown because of the uncertainty and non-measurability of $r_i$, but its first and second-order moments are estimable. Expected abatement is $E_i = \hat{r}_i - \mu_i$ and the variance is $c_i = \sigma_i^2$.

Market rules must be unambiguously defined for a point-nonpoint market to succeed. Existing water quality trading markets like the Lower Boise River program and New York City Watershed Offsets Pilot Program performed poorly because of uncertainty in the regulatory environment (Morgan and Wolverton, 2005; Breetz et al., 2004). Hence, the NPS permit should not be based on the unknown abatement $a_i$. Instead, we define $i$’s permit supply as the multi-attribute good $\bar{a}_i(E_i(\mu_i), c_i(\sigma_i^2))$. The permit’s two attributes are deterministically and independently controllable, as discussed in Section 2, by manipulation of the parameters of the implemented BMPs.

However, defining the permit as a multi-attribute good raises the problem of divisibility. How does one define one-sixth of a house, the quintessential example of a multi-attribute good? In our problem context, what does a PS buy when he trades with an NPS? Must he buy the nonpoint’s entire permit supply much like a prospective buyer purchases an entire house (and not, say, a single bedroom)? Are permits to be separable by technology? Can the PS buy only those permits generated by the riparian buffer, but not those generated by the nutrient management plan? This is analogous to the home buyer purchasing those parts of the property built by Tom, but not those by Harry. Clearly such transactions would be illogical and needlessly complicated. Instead, much like multiple parties can jointly purchase a property, each

---

abatement is $e_j^p - e_j$. However, $j$ is only allotted permits equivalent to $\delta_j - e_j$. The results in this paper hold irrespective of which type of abatement is considered. We choose the former for notational convenience.
obtaining an interest in a share of the whole, we predicate trades upon the nonpoint source’s total abatement.

The PS defrays a share of the NPS’s total investment in abatement and is entitled to a commensurate share of the resultant uncertain abatement. For example, consider the Pennsylvanian farmer $i$ implementing a riparian buffer and a nutrient management plan. The attributes in $i$’s permit supply $\bar{a}_i(E_i, c_i)$ are determined by the modeled reduction of runoff as a consequence of these technologies. A PS $j$, perhaps a wastewater treatment plant, seeks to trade with $i$. We allow $j$ to buy a $\gamma_{ij} \in [0, 1]$ proportion or share of $\bar{a}_i(E_i, c_i)$ where $\gamma_{ij} = 0$ implies that $j$ does not trade with $i$ and $\gamma_{ij} = 1$ implies that $j$ buys $i$’s entire permit supply. This is equivalent to $j$ investing ex post in a $\gamma_{ij}$ share of $i$’s BMPs. From $j$’s point of view, buying shares from multiple nonpoint sources makes sense because it spreads the risk that it will violate its allowance. Since the market is competitive, each $i \in I$ sells its entire permit supply, i.e. $\sum_i \gamma_{ij} = 1 \forall i$.

Prior to market entry NPS $i$ produced emissions at its unconstrained profit maximizing level $\pi_i^0$. After entering the market $i$ implements some BMPs. It produces and sells $\bar{a}_i(E_i, c_i)$ permits and its restricted profit function is $\pi_i(\mu_i, \sigma_i^2)$. The cost of generating permits is $\pi_i^0 - \pi_i(\mu_i, \sigma_i^2)$. Upon selling $\bar{a}_i(E_i, c_i)$ permits $i$ earns $q_i(E_i, c_i)$ in revenue. Nonpoint sources do not buy permits. Total profits from permit production and trade are $\Pi_i = q_i \left( E_i(\mu_i), c_i(\sigma_i^2) \right) - \left[ \pi_i^0 - \pi_i(\mu_i, \sigma_i^2) \right]$. NPS $i$ faces an unconstrained optimization problem and has 2 control variables, $\{\mu_i, \sigma_i^2\}$. The necessary conditions for optimization are

$$\frac{\partial \pi_i}{\partial \mu_i} = -\frac{\partial q_i}{\partial \mu_i}$$

(4)

$$\frac{\partial \pi_i}{\partial \sigma_i^2} = -\frac{\partial q_i}{\partial \sigma_i^2}$$

(5)
Since $\partial \pi_i / \partial \mu_i \geq 0$ and $\partial \pi_i / \partial \sigma_i^2 \geq 0$, (4) and (5) indicate that optimally, sales revenue $q_i$ should decrease in $\mu_i$ and $\sigma_i^2$ at the profit maximum. Or equivalently, $i$’s revenues from permit sale increase in the mean $E_i$ and decrease in the variance $c_i$ of its permits.

PS $j$ has an emissions allowance $\hat{e}_j$. Under autarky, it meets its allowance by choosing some emissions level $e_j \leq \hat{e}_j$. Under trade, $j$ augments its allowance by buying permits from other polluters. Let $\gamma_{kj} \in [0,1]$ be the proportion of $k$’s permits that $j$ buys where $k \in J$ is another PS. $\gamma_{kj}a_k$ is the number of permits that $j$ buys from $k$ and $\sum_{k \neq j} \gamma_{kj}a_k$ are $j$’s purchases from all other point sources. Analogously, its permit purchases from all nonpoint sources is $\sum_i \gamma_{ij}\bar{a}_i(E_i,c_i)$. In addition, it creates $a_j$ of its own permits, of which $\gamma_{jk}a_j$ permits are sold to $k \in J$, $\sum_{k \neq j} \gamma_{jk}a_j$ permits are sold to all other point sources and $\gamma_{jj}a_j$ permits remain unsold. After trade, $j$ is constrained to choose some $e_j \leq \hat{e}_j + \sum_i \gamma_{ij}\bar{a}_i - \sum_{k \neq j} \gamma_{jk}a_j + \sum_{k \neq j} \gamma_{kj}a_k$. Using the tautology that $\sum_k \gamma_{jk} = \sum_{k \neq j} \gamma_{jk} + \gamma_{jj} \equiv 1$, this constraint can be rewritten as $\sum_k \gamma_{kj}a_k + \sum_i \gamma_{ij}\bar{a}_i \geq 0$. Given how $\bar{a}_i$ is defined, the true difference between the allowance and emissions is $\sum_k \gamma_{kj}a_k + \sum_i \gamma_{ij}a_i$. The mean difference is $\sum_k \gamma_{kj}a_k + \sum_i \gamma_{ij}E_i$ and the variance is $\sum_i \gamma_{ij}^2c_i^2$.

PS $j$’s pollution liability is stochastic because of its purchases from nonpoint sources. Analogous to the EA’s problem we proceed under a safety first framework and constrain this liability probabilistically. Let $\beta \in [0,1]$ be the index measuring the EA’s tolerance for any PS exceeding its permit-augmented allowance. It is not necessary that $\beta = \alpha$, the tolerance for violating the aggregate safety first constraint. The EA has different preferences over aggregate and individual emissions. Intuitively this makes sense: there is no reason for the EA to care as much about the likelihood of an individual polluter violating its cap as it does about the aggregate cap being violated. An individual cap violation is not necessarily in conflict with
compliance with the safety first constraint. Under-abatement by some polluters may be accompanied by commensurate over-abatement by others. Fundamentally, the choice of $\beta$ must be consistent with the EA’s goal of meeting the safety first constraint at least cost. As is shown later, setting $\beta = \alpha$ is generally not consistent with this goal. Controlling $j$’s emissions in a safety first framework requires that $\Pr(\sum_k y_{kj} a_k + \sum_i y_{ij} a_i \geq 0) \geq 1 - \beta$. Applying Chebychev’s inequality, $j$’s allowance constraint is $\sum_k y_{kj} a_k + \sum_i y_{ij} E_i - \sqrt{\sum_i y_{ij}^2 c_i / \beta} \geq 0$.

Substituting for $\mu_i, \sigma_i^2$ and $e_j$ the constraint may be rewritten in terms of emissions as

$$\sum_j y_{kj} e_k + \sum_i y_{ij} \mu_i + \sqrt{\sum_i y_{ij}^2 \sigma_i^2 / \beta} \leq \sum_k y_{kj} a_k + \sum_i y_{ij} \hat{a}_i$$

(6)

where $\sum_k y_{kj} \hat{a}_k + \sum_i y_{ij} \hat{a}_i$ is $j$’s allowance after trade. $y_{kj} e_k$ is the share of $k$’s emissions and $y_{ij} \mu_i$ and $y_{ij}^2 \sigma_i^2$ are the mean and variance of the share of $i$’s emissions that $j$ is liable for. $\sum_j y_{kj} e_k, \sum_i y_{ij} \mu_i$ and $V_j \equiv \sum_i y_{ij}^2 \sigma_i^2$ are the total PS emissions, and the mean and variance of total NPS emissions that $j$ is liable for. It can vary $e_j, \{y_{kj}\}_{k \in j}$ and $\{y_{ij}\}_{i \in j}$, but not the attributes of the emissions of others. The allowance constraint (6) and the safety first aggregate constraint (1) are identical except for the $\gamma$ weights in the former, which distort the relationship between variables away from what is required to satisfy (1). Otherwise the allowance constraint is interpreted analogously.

Note that a PS is not liable for actual emissions. It is compliant if it buys enough permits such that it satisfies (6) ex ante. It is not culpable for adverse random events that increase ex post emissions beyond $\sum_k y_{kj} \hat{a}_k + \sum_i y_{ij} \hat{a}_i$. This is an attractive design feature for two reasons. First, it is not fair to penalize the PS for events beyond its control. Second, the non-measurability of $a_i$ makes proving culpability problematic. This feature does not exist in real world WQT markets,
where the NPS abatement becomes part of the PS NPDES permit. The PS then becomes liable for the ex post emissions violation of contracted NPS.\textsuperscript{8}

Let PS \( j \)'s unconstrained production profits be \( \pi_j^0 \) and restricted profits be \( \pi_j \left( a_j(e_j) \right) \). The cost of permit generation is \( \pi_i^0 - \pi_j(a_j) \). Let \( p \) be the unit market price of a PS permit. Since the market is competitive, it is exogenous. Total payment to other point sources for permits purchased is \( p \sum_{k \neq j} \gamma_{kj} a_k \). The payment received from permit sales is \( p \sum_{k \neq j} \gamma_{jk} a_j \). It follows that net payments to other point sources are \( p \left[ \sum_k \gamma_{kj} a_k - a_j \right] \). PS \( j \)'s total payments to nonpoint sources are \( \sum_i \gamma_{ij} q_i(E_i, c_i) \). Summing, \( j \)'s total profits from production and trade of permits are

\[
\Pi_j = p \left[ a_j - \sum_k \gamma_{kj} a_k \right] - \left[ \pi_j^0 - \pi_j(a_j) \right] - \sum_i \gamma_{ij} q_i(E_i, c_i). \tag{6}
\]

It maximizes \( \Pi_j \) while satisfying (6) by controlling permit generation \( a_j \), and its portfolio of other polluters' permits, \( \{ \gamma_{kj}, \gamma_{ij} \}_{i,k} \). The necessary conditions for profit maximization are (7) and (8).

\[
\frac{\partial \pi_j}{\partial e_j} = p \tag{7}
\]

\[
q_i^* = p \left( E_i - \gamma_{ij} c_i / \sqrt{BV_j} \right) \forall i \in I \tag{8}
\]

The first necessary condition (7) indicates that \( j \) optimizes emissions (or abatement) by equating marginal production profits to the market price of PS permits. In (8) \( q_i^* \) is \( i \)'s revenue from supplying \( a_i \) when \( j \) maximizes profits. Consider some \( k \in J, k \neq j \). The profit maximizing necessary conditions for \( k \) are also (7) and (8). When both \( k \) and \( j \) maximize profits simultaneously then \( q_i^* = p \left( E_i - \gamma_{ij} c_i / \sqrt{BV_j} \right) = p \left( E_i - \gamma_{ik} c_i / \sqrt{BV_k} \right) \) from (8). This is true only when \( \gamma_{ij}/\gamma_{ik} = \sqrt{V_j}/\sqrt{V_k} \). Since (8) holds for all nonpoint sources, it follows that \( \gamma_{ij}/\gamma_{ik} = \)

\textsuperscript{8}The difference in PS liability between the MANA market and real world markets is in when the liability ends. In the MANA market, PS liability ends with correct implementation. In real world markets, liability ends later, after the emissions measurement period ends.
\(\sqrt{V_j}/\sqrt{V_k}\) for any \(l \in I, l \neq i\). Hence, for an arbitrary pair of point sources to simultaneously maximize profits, equation (9) is necessary:

\[
\frac{\gamma_{ij}}{\gamma_{ik}} = \frac{\gamma_{ij}}{\gamma_{ik}} = \frac{\sqrt{V_j}}{\sqrt{V_k}} \quad \forall \ i, j, k, l \in I \times J
\]  

(9)

The first equality in (9) indicates that PS j’s share in NPS i’s abatement relative to PS k’s share is equal to its share in any other NPS l’s abatement relative to k’s share in the same. If j’s share in i is twice as much as k’s then its share in l will also be twice as much. The second equality indicates that this ratio is equal to the ratio between the total standard deviation of their emissions liabilities (or, equivalently, the total standard deviation of their purchased permits). The total standard deviation is an index or proxy of the risk that a PS faces that it will exceed its permit-augmented allowance. This risk comes entirely from its portfolio of NPS permits. Hence equation (9) indicates that all pairs of point sources buy NPS permits in a fixed ratio equal to the ratio of their total portfolio risks.

A necessary condition for fixed ratios to exist for all pairs is for each PS to buy the same share from all nonpoint sources, i.e. \(\gamma_{ij} = \gamma_{j} \ \forall \ i \in I, j \in J\). The proof is as follows: Let \(\gamma_{ij} = b_{j}\gamma_{ij}\) and \(\gamma_{ik} = b_{k}\gamma_{ik}\) for any pair of point sources \(j, k \in J\) and any pair of nonpoint sources \(i, l \in I\). For (9) to hold, \(b_{j} = b_{k} = b\). Since the market is competitive and all NPS sell all their permits, \(\gamma_{ik} = 1 - \sum_{j \neq k} \gamma_{ij}\) and \(\gamma_{lk} = 1 - \sum_{j \neq k} \gamma_{lj}\). But since \(\gamma_{ij} = b\gamma_{ij} \ \forall \ j\) it follows that \(\gamma_{ik} = 1 - b\sum_{j \neq k} \gamma_{ij}\). Hence, the condition \((1 - \sum_{j} \gamma_{ij})/\gamma_{ij} = (1 - b\sum_{j} \gamma_{ij})/b\gamma_{ij}\) must hold for (9) to hold. But this condition only holds when \(b = 1\). This implies that \(\gamma_{1j} = \cdots = \gamma_{lj} = \gamma_{j}\) for all \(j \in J\). It also follows that \(\sum_{j} \gamma_{j} = 1\). Any PS j buys a \(\gamma_{j}\) share in the abatements of all nonpoint sources.
**Result 1.** In a competitive market point sources control their exposure to the risk of exceeding their allowances by buying identical shares in all nonpoint source permit-generating projects. This behavior may be described as **Equi-Proportional Risk Sharing** across the different nonpoint permits.

By Result 1, j’s allowance constraint amends from (6) to \( \sum_k y_{kj} a_k + y_j (\sum_i E_i - \sqrt{V/\beta}) \geq 0 \), but its profit function remains unchanged. The necessary conditions for j’s profit maximization transform to (7) and (10).

\[
\sum_i q_i^* = p \left( \sum_i E_i (\mu_i) - \frac{V}{\sqrt{\beta V}} \right)
\]

(10)

The necessary condition (10) identifies the total payment to all nonpoint sources. \( \sum_i q_i^* \) is the total payment from all j to all i. On the RHS \( \sum_i E_i \) and \( V = \sum_i c_i \) are the mean and variance of aggregated NPS abatements. The total payment \( \sum_i q_i^* \) is additively separable into functions of \( E_i \) and \( c_i \) for all \( i \in I \). It marginally increases in \( E_i \) by \( p \). Since \( p \) is also the marginal cost of \( a_j \) (see (7)), \( E_i \) and \( a_j \) are perfect substitutes at the point source’s profit maximizing margin. The marginal payment for \( c_i \) is \(-p/2\sqrt{\beta V}\). NPS i is penalized \( p/2\sqrt{\beta V} \) for increasing \( c_i \) by one unit since an increase in \( c_i \) increases the probability of constraint violation. This penalty is the PS permit price inflated by \( 1/2\sqrt{\beta V} \), the marginal uncertainty penalty for exceeding j’s allowance. The penalty is analogous to \( 1/2\sqrt{\alpha V} \), the marginal uncertainty penalty for violating the aggregate safety first constraint. The penalty does not vary across point sources. All violators pay the same penalty irrespective of their individual characteristics and levels of violation.

Note that \( q_i^* \) is not explicitly specified in (10). As long as (10) holds, \( q_i^* \) could be described by any function. Each NPS could face a different payment function. However, given (10) a simple functional form \( q_i^* = p(E_i(\mu_i) - c_i/\sqrt{\beta V}) \forall i \) presents itself. It is easy to see that
summing over \( q_i^* \) yields (10) and hence maximizes \( j \)'s profits given equi-proportional risk sharing. The inferences on the separability of \( \sum_i q_i^* \) hold for \( q_i^* \). The payment function \( q_i^* \) imposes conditions on the characteristics of the permits \( \tilde{a}_i(E_i, c_i) \). NPS \( i \) gets a positive payment if and only if \( c_i/E_i \leq \sqrt{\beta V} \) where \( c_i/E_i \) is \( i \)'s variance-to-mean ratio. \( \sqrt{\beta V} \) is half the inverse of \( j \)'s marginal uncertainty penalty and is a proportionate index of the EA's tolerance for any PS exceeding its allowance.

Equations (7) and (10) are necessary conditions only when \( j \) trades with both point and nonpoint sources. If \( j \) only traded with other point sources then (7) is the only relevant necessary condition. However, such behavior is sub-optimal when NPS permit generation is cheaper than PS permit generation, as is typically the case. If \( j \) traded with only nonpoint sources then (7) and (10) are necessary for profit maximization under the caveat that \( p \) exists. The existence of \( p \) is predicated upon the existence of trade among point sources. If such trade did not exist then \( p \) may be set by reference to prices in other tradable permit markets. Other scenarios, such as \( j \) not trading at all or only selling permits, are sub-optimal when the allowance constraint is binding and PS abatement is expensive.

### 3.1 The Market Equilibrium

The MANA market equilibrium is characterized by all point and nonpoint sources simultaneously maximizing their profits, i.e. equations (4), (5), (7) and (10) must hold simultaneously. It follows that the MANA market competitive equilibrium is characterized by (11) below.

\[
p = \frac{\partial \pi_i}{\partial e_j} = \frac{\partial \pi_i}{\partial \mu_i} = \frac{\partial \pi_i}{\partial \sigma^2_i} \frac{\sqrt{\beta V}}{1 - \sigma^2_i/2V} \quad \forall \ i \in I, j \in J
\]

At equilibrium, marginal profits with respect to measured PS profits and mean NPS profits are equal to the PS permit price, \( p \). NPS adjust their marginal profits with respect to variance
upwards by $\sqrt{\beta V}$, an index of the EA’s tolerance for any PS exceeding his allowance. This index of individual tolerance exceedence is itself inflated upward commensurate to $i$’s share of the aggregate emissions variance $\sigma_i^2 / V$. When $\sqrt{\beta V}$ decreases, the nonpoint source optimizes by reducing its emissions variance, which increases marginal profits, thereby restoring (11). By concavity, high variance polluters must reduce their variance by a greater amount than low variance polluters. This inference follows from the observation that under concavity the marginal response to changes in variance is smaller when the variance is high.

In a competitive market $\sigma_i^2 / 2V \rightarrow 0$, which implies that $\sqrt{\beta V}$ is not inflated upward by $i$’s variance share and $\partial \pi_i / \partial \sigma_i^2 \rightarrow p / \sqrt{\beta V}$. Since $\sqrt{\beta V}$ is identical across $i$, $\partial \pi_i / \partial \sigma_i^2$ is equal across NPS. The concavity of $i$’s profits implies that NPS $i$ chooses a high $\sigma_i^2$ when tolerance for individual exceedences is high and vice versa. The intuition might be as follows: consider some $i$ that keeps $\sigma_i^2$ low when $\sqrt{\beta V}$ is high. By reducing $\sigma_i^2$ marginally it earns $p / \sqrt{\beta V}$, which is small. But concavity of $\pi_i$ in $\sigma_i^2$ implies that the marginal cost of reducing $\sigma_i^2$ is high when $\sigma_i^2$ is low. An assessment of costs and benefits would convince $i$ to keep $\sigma_i^2$ high when $\sqrt{\beta V}$ is high.

**Result 2.** The MANA market is in competitive equilibrium when the marginality conditions (4) and (5) hold for all nonpoint sources, contingent on each being paid $q_i^* = p\left(E_i - \sigma_i^2 / \sqrt{\beta V}\right)$, and the marginality condition (7) holds for all point sources. These conditions are summarized in (11).

The pertinent issue for the EA is whether conditions can be imposed on the market such that the safety first constraint (1) is met at least cost at the market equilibrium. In comparing the conditions define the MANA market equilibrium (11) and the EA’s optimum (3), we see that the EA and polluters value $e_j$ and $\mu_i$ identically, but that their valuations of $\sigma_i^2$ differ.
valuations are identical only when $2\sqrt{\alpha} = \sqrt{\beta}/(1 - \sigma_i^2/2V)$. Since the EA has the freedom to choose $\beta$, it can always ensure that this relationship holds. It follows that, in theory, the EA can always ensure that the MANA market equilibrium meets the safety first constraint at least cost by varying its tolerance for individual allowance exceedences. In practice, optimally adjusting $\beta$ is not straightforward. To set $\beta$ optimally, the EA must know $\sigma_i^2 \ \forall \ i$ beforehand, i.e. in the Pennsylvania market, the EA must know of the BMPs that $i$ implements, as well as how these BMPs are implemented. But since $i$ only chooses its BMPs (and thus $\sigma_i^2$) after knowing $\beta$ a chicken-and-egg situation develops.

The distortions that might result from this paradox will be negligible in a competitive market. There will be many atomistic NPS in such a market, which implies that the emissions variance of any single NPS will be vanishingly small when compared to the aggregate, i.e. $\sigma_i^2/2V \to 0$ for all $i$. It follows that, in a competitive market equilibrium, the EA and all polluters will have identical valuations of $\sigma_i^2$ and the equilibrium will satisfy the safety first constraint at least cost when $\beta \to 4\alpha$. If the tolerance probability at the aggregate level is 5 percent then the EA can ensure least cost compliance with the safety first constraint by setting the tolerance probability for individual allowance exceedence at 20 percent. The EA’s different valuations of aggregate and individual emissions variances also explain why $\beta = \alpha$ is almost universally suboptimal. Intuitively, the only situation where $\beta = \alpha$ is optimal is when individual and aggregate emissions variances are identical, i.e. when there is a single NPS in the market. This intuition is borne out by noticing that the only scenario where the two equalities $\beta = \alpha$ and $2\sqrt{\alpha} = \sqrt{\beta}/(1 - \sigma_i^2/2V)$ hold is where $\sigma_i^2 = V$.

**Result 3.** The equilibrium in the MANA market will meet the safety first outcome at least cost under a single condition: that the ratio between the tolerances for individual allowance and
safety first constraint violation be chosen such that $\sqrt{\beta/\alpha} = 2(1 - \sigma_i^2/2V)$. When the market is competitive and $\sigma_i^2/V \to 0$, then the equilibrium is optimum when $\beta/\alpha \to 4$.

4 The Trading Ratio Market

Extant water quality trading markets do not use a multi-attribute definition of the NPS permit. Instead the permit is defined as the expected value, $E_i = \hat{r}_i - \mu_i$, of the true reduction in emissions below the allowance, $a_i = \hat{r}_i - r_i$. Unlike in the MANA market, $i$’s permit is a first order description of $a_i$. This description prevents transmission of information about $i$’s emissions variance $\sigma_i^2$ to the marketplace. Also, PS allowances in extant water quality trading markets are not probabilistic, which implies that $\sigma_i^2$ will not directly affect trading decisions.\(^9\)

Instead of using the market mechanism to control $\sigma_i^2$, an ad hoc adjustment called the ‘uncertainty trading ratio’ is used. The trading ratio is defined as the units of mean NPS abatements required to allow a unit increase in PS emissions (Shortle, 1990). It is calculated by the EA and applied to all trades with NPS. Proponents of the trading ratio argue that when correctly designed it addresses the risk associated with uncertain nonpoint emissions. It converts expected NPS emissions into perfect substitutes for actual PS emissions, so that in effect apples will trade for apples. The optimal design of trading ratios is examined in several papers including Shortle (1987) and Hennessy and Feng (2008). However, these papers analyze a situation where a single PS trades with a single NPS and the EA’s objective is to limit some social or environmental damage function. Our model more closely simulates a real world environment in that we assume multiple polluters of both types and that the EA’s objective is in

\(^9\)Point sources do face legal liability if the water quality based effluent limit (WQBEL) – the tradable part of its ERC – is violated. That is, if the instream condition that the PS must meet is not met, the PS is liable. Instream water quality is a function of loadings stochasticity, but under the trading ratio framework, there is no way to explicitly incorporate this stochasticity into the decision making framework.
limiting emissions below some exogenously chosen limit dependent on a TMDL or analogous regulation.

Let $t_{ij}$ be the trading ratio applied to a trade between $i$ and $j$. Analogous to Horan (2001) the emissions allowance for $j$ is $\sum_k y_k a_k + \sum_i (y_{ij}/t_{ij})E_i(\mu_i) \geq 0$. PS $j$ maximizes $\Pi_j = p [a_j - \sum_k y_{kj} a_k] - [\pi^0_j - \pi_j(a_j)] - \sum_{i} y_{ij} q^t_i(E_i)$ subject to this allowance by choosing $E_i, \{y_{kj}\}_{k\in j}$ and $\{y_{ij}\}_{i\in I}$. $\Pi_j$ is identical in both markets except that the NPS payment function changes. NPS are no longer paid for controlling $\sigma^2_i$. The necessary conditions are (7) and (12).

$$q^t_i = p \frac{E_i}{t_{ij}} \quad \forall \ i \in I$$

(12)

where $q^t_i$ is NPS $i$’s sales revenue associated with $j$’s profit maximum. PS $j$ pays NPS $i y_{ij} q^t_i$ for supplying $y_{ij} E_i$ permits. By (12) and the definition of $E_i$, these payments decrease as $t_{ij}$ or $\mu_i$ increase. NPS with high mean emissions are paid less, which is consistent with a policy aimed at reducing mean emissions. Since (12) holds for all PS it follows that $q^t_i = pE_i/t_{ij} = pE_i/t_{ik}$ and $t_{ij} = t_{ik} = t_i$ for all $j, k \in J$. The trading ratio varies across nonpoint sources but not across individual trades at equilibrium. This result differs from previous results where the trading ratio is trade-specific, affected as it is by PS and NPS characteristics (e.g. Horan, 2001). The difference is a consequence of scale. Whereas earlier models focused on a single PS and a single NPS, we focus on the entire market.

In the trading ratio market, like in the MANA market, NPS $i$ has a permit generating cost, earns revenues from permit sales and does not face a loadings constraint. Total profits, $\Pi_i = pE_i(\mu_i)/t_i - [\pi^0_i - \pi_i(\mu_i)]$, are maximized by optimally choosing $\mu_i$. The necessary condition is
By (13) the equilibrium trading ratio is \( t_i^* = \frac{p}{t_i} \). Since \( \pi_i \) is increasing and concave in \( \mu_i \) it follows that \( t_i^* \geq 0 \) and increases in \( \mu_i \). Substituting \( t_i^* \) into (12) yields \( q_{i}^{i*} = (\hat{r}_i - \mu_i) \frac{\partial \pi_i}{\partial \mu_i} = -E_i \frac{\partial \pi_i}{\partial E_i} \) where \( q_{i}^{i*} \) maximizes both PS and NPS profits. \( \frac{\partial \pi_i}{\partial \mu_i} \) identifies the optimal unit price of \( i \)'s abatement. Sales revenues approaching their maximum as \( \mu_i \to 0 \), monotonically decrease in \( \mu_i \) and approach zero as \( \mu_i \to \hat{r}_i \). Both PS and NPS have strong preferences for permits with low \( \mu_i \). At the market equilibrium all polluters are profit maximizers and (7), (12) and (13) hold simultaneously for all \( i \) and \( j \). The necessary condition is shown in (14). PS and NPS have identical marginal costs of permit generation after adjusting for the trading ratio.

\[
\frac{\partial \pi_j}{\partial e_j} = t_i^* \frac{\partial \pi_i}{\partial \mu_i} \quad \forall i, j \in I \times J
\]  

(14)

In general, the trading ratio market is never safety first compliant at least cost because it incentivizes control of fewer variables than necessary for compliance. The \( \sigma_i^2 \)s, each critical to safety first compliance, are not controlled. Indeed, given the market rules, it is only through happenstance that the equilibrium might be optimal. For this to come to pass, it is necessary that controlling \( \mu_i \) \( \forall i \) and \( e_j \) \( \forall j \) results in optimal control of \( \sigma_i^2 \) \( \forall i \). Given independence between NPS, this requires \( \sigma_i^2 \) as a deterministic function of \( \mu_i \) or \( \sigma_i^2 = \sigma_i^2(\mu_i) \).

This assumption is the exact opposite of the independence assumption in the MANA market analysis. Whereas \( \mu_i \) and \( \sigma_i^2 \) were assumed to vary freely in the MANA market, here they have a specific non-stochastic relationship. It validity depends on two conditions. First, \( i \)'s emissions must follow a probability distribution where \( \sigma_i^2 = \sigma_i^2(\mu_i) \). Candidates include the binomial, exponential, hypergeometric and Poisson distributions. Second, this distribution must be identifiable given the current state of the science. The feasibility of the first condition for all
nonpoint sources is unknown. Satisfying the second condition is problematic, given the uncertainty associated with current models of BMP effectiveness (e.g. Osborne and Kovacic, 1993; van Dyke et al., 1999). Typically many distributions can be reasonably fitted to a given data set and each will define a different relationship between $\mu_i$ and $\sigma_i^2$. Choosing the best distribution will be necessarily arbitrary and often justified by nothing more than common practice, which in turn implies that $\sigma_i^2(\mu_i)$ will be arbitrarily assigned. Disregarding these concerns, we define the necessary conditions for compliance with the safety first constraint (1) at least cost in the trading ratio market in (15).

$$\frac{\partial \pi_j}{\partial \varepsilon_i} = \frac{\partial \pi_i}{\partial \mu_i} \left(1 + \frac{1}{2\sqrt{\alpha N}} \frac{\partial \sigma_i^2}{\partial \mu_i}\right)^{-1} \forall \ i \in I, j \in J \quad (15)$$

The marginal response of $\sigma_i^2$ to $\mu_i$, $\partial \sigma_i^2/\partial \mu_i$, is $\mu_i$’s marginal risk and may be positive or negative. Consider both cases: let $\partial \sigma_i^2/\partial \mu_i \geq 0$ in case 1 and $\partial \sigma_i^2/\partial \mu_i < 0$ in case 2. For (15) to hold it must be that $\partial \pi_i/\partial \mu_i$ is higher in case 1 than in case 2. By concavity, it follows that $\mu_i$ is higher in case 2. NPS i’s permits will embody emissions with low $\mu_i$ and $\sigma_i^2$ in case 1 and emissions with high $\mu_i$ but low $\sigma_i^2$ in case 2. The permits in case 1 are better from the EA’s point of view because they will weaken (1) to a greater extent. The EA prefers case 1 where $\partial \sigma_i^2/\partial \mu_i \geq 0 \forall i$.

The trading ratio market will satisfy the safety first constraint at least cost when the necessary conditions for its competitive equilibrium (14) are identical to those for the safety first optimum when $\sigma_i^2 = \sigma_i^2(\mu_i)$ (15). This requires that

$$\frac{1}{t_i^{**}} = 1 + \frac{1}{2\sqrt{\alpha N}} \frac{\partial \sigma_i^2}{\partial \mu_i} \forall i \in I, \quad (16)$$

Equation (16) indicates that $t_i^{**} \leq 1$ in case 1 and $t_i^{**} > 1$ in case 2. The lower the trading ratio the more favorable the terms of trade for the NPS and, by (12), the higher are the revenues.
from permit sale. In case 1 since $t_i^{**}$ is low, $\partial \pi_i / \partial \mu_i$ is high [see (13)] and, by concavity, $\mu_i$ is low. Since $\partial \sigma_i^2 / \partial \mu_i \geq 0$, $\sigma_i^2$ is also low. Since both reductions are aligned with the EA’s preferences, it favors NPS with $\partial \sigma_i^2 / \partial \mu_i \geq 0$ by giving them advantageous terms of trade. Conversely, the EA gives disadvantageous terms of trade to NPS under case 2 by setting $t_i^{**} > 1$. Such BMPs cannot reduce both mean emissions and the variance simultaneously and are less appealing to the EA.

Choosing the optimal trading ratio requires perfect information and foresight. The EA needs ex ante knowledge of $\sigma_i^2(\mu_i) \forall i$, as well as the ability to perfectly forecast the market equilibrium. This is impossible in the real world. Indeed, if perfect information and foresight existed, then the market would be redundant. The EA could directly assign optimal abatement projects across polluters and then monitor their behavior to ensure that they behaved optimally. It follows that the real world trading ratio market will fail to comply with the safety first constraint at least cost.

**Result 4.** Equilibrium in the trading ratio market satisfies the safety first pollution constraint at least cost under three conditions. First, the EA must have perfect information. Second, the variances of all nonpoint emissions must be deterministic functions of the mean. Finally, $t_i^{**}$ must be defined according to (16).

In existing markets the trading ratio exceeds one, which implies that these markets function under the implicit assumption that case 2 is valid (in addition to the other assumptions about perfect information and the existence and identifiability of $\sigma_i^2(\mu_i)$). These markets are always suboptimal when, in truth, case 1 holds. Often $t_i^{**} = 2 \forall i$ (Breetz et al., 2004), which is optimal only when case 2 is true and $\partial \sigma_i^2 / \partial \mu_i = -\sqrt{\alpha V}$ for all $i$, where $\sqrt{\alpha V}$ reflects societal tolerance for violating the aggregate safety first constraint. This factor will be high when $V$ is
high, in turn implying that $\sigma_i^2$’s optimal marginal response to $\mu_i$ must be high. The feasibility of such a response is beyond the scope of this paper, but is unlikely. When the response is weaker, then $t_i^{**}$ will be closer to one than two. In all likelihood a uniform trading ratio of two is sub-optimal.

5 Discussion and Conclusions

Trading ratio markets are the current state-of-the-art in point-nonpoint trading. These markets are unsatisfactory when it comes to controlling the risk that ex post emissions exceed pre-specified emissions caps. There are many reasons for this. First, optimality in the trading ratio market requires a deterministic relationship between the mean and variance of emissions, both of which are empirically determined. The existence and identification of this relationship are both unlikely. Second, the requirement that the EA set the trading ratio for all nonpoint sources places a heavy information burden on it. Indeed, the common practice is to use a uniform trading ratio of two, which is only optimal if the marginal response of variance to mean is highly negative for all NPS. This information burden is separate from the other tasks of certifying NPS abatements for permit generation and providing permit information to the marketplace.

We propose the MANA market as a workable alternative to the trading ratio market and prove that it controls NPS pollution in a safety first framework at least cost. The two markets are identical in their setup, but the MANA market has features that improve its performance vis-à-vis the trading ratio market. First, the MANA market does not require a deterministic relationship between the mean and variance of nonpoint emissions. Indeed, it assumes the opposite: that these variables are independent. Independence, as discussed in Section 2, is reasonable in the context of real world point-nonpoint programs. Second, the MANA market is
optimal when loads are correlated because of the mechanisms by which BMPs reduce emissions, i.e. emissions distributions get truncated. Third, the EA’s information burden is much smaller than in the trading ratio market. The EA does not need to intervene and define a trading ratio for every NPS. Indeed, as long as the market is competitive, the EA need only set the tolerance for individual allowance exceedences at four times the tolerance for violation of the safety first constraint. In a real world market, like the Pennsylvanian programs, the EA’s role will be purely informational: it is only required to provide information on $\mu_i$ and $\sigma_i^2$ for all $i \in I$.

Numerous extensions to the model are possible. An interesting possibility is to conduct a series of economic experiments to test the robustness of the results to violations in underlying assumptions. The assumption of competitiveness in the market can be removed. Real world markets are oligopsonistic with few large PS and many small NPS. Other extensions could deal with how the baselines used to define NPS permits affect the supply and price of these permits and whether these changes affect market outcomes. These extensions are relevant to policy design because they represent an intermediate step in the transfer of policy ideas from the academic to the real world.
Appendix

Consider a line of six farms that lie along a streambank and are indexed $A, B, \ldots, F$. $A$ and $F$ have one immediate neighbor and the others have two. Let the emissions variances of these farms be \{0.5, 1.0, \ldots, 3.0\}. Each farm is correlated with all others and the correlation decreases with distance. Correlation with immediate neighbors is 0.6, correlation with neighbors once removed is 0.5 and so on. The emissions are assumed to be distributed multivariate normal.

We simulate how the emissions variances and correlations attenuate as the distribution is truncated by BMPs. The simulation is conducted in R, and the Gibbs sampler was used to draw from the truncated distributions (see Wilhelm and Manjunath (2010) for details). In Figure 1, we plot the emissions correlations of each farm with its neighbors for different levels of BMP effectiveness. We assume that all farms truncate equally, but results and inferences do not change if this were not true. As truncation increases from zero to 75 percent of the full distribution, correlation falls towards zero for all farm pairs. Beyond a certain level of technological efficiency, correlations attenuate to such a degree that the emissions become approximately independent. Hence, emissions might legitimately be considered independent for highly effective abatement technologies like riparian buffers.

But what about less effective technologies like nutrient management plans? Figure 1 indicates that correlations might be fairly high even after a 40 percent truncation. In Figure 2 variances and covariances associated with each farm are plotted in separate panels. Two observations can be made. First, all variances and covariances fall towards zero as the distribution gets truncated. Second, all covariances reach zero at lower levels of truncation than the variances. It follows that the effectiveness of a technology in contributing to compliance with the safety first constraint can be gauged merely by focusing on how it affects the variance.
of emissions. The EA can assume that NPS emissions are independent for the purpose of analysis.
Figure 1: BMPs and their effect on Correlation between Nonpoint Source Emissions
Figure 2: BMPs and their effect on the Variance and Covariances of Nonpoint Source Emissions
References


J. Dales, 1968, “Pollution, Property and Prices,” University of Toronto Press, Toronto.


R. Merz, G. Blöschl 2009, A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria, *Water Resources Research* 45, W01405.


List of FCN Working Papers

2012


2011


2010


2009


2008


FCN Working Papers are free of charge. They can mostly be downloaded in pdf format from the FCN / E.ON ERC Website (www.eonerc.rwth-aachen.de/fcn) and the SSRN Website (www.ssrn.com), respectively. Alternatively, they may also be ordered as hardcopies from Ms Sabine Schill (Phone: +49 (0) 241-80 49820, E-mail: post_fcn@eonerc.rwth-aachen.de), RWTH Aachen University, Institute for Future Energy Consumer Needs and Behavior (FCN), Chair of Energy Economics and Management (Prof. Dr. Reinhard Madlener), Mathieustrasse 6, 52074 Aachen, Germany.