Breakdown of the Equi-Marginal Principle in Permit Markets
Involving Multiple Pollutants and Exogenous Caps

Abstract Interest in expanding market-based approaches involving intra-pollutant trading (trading “like” pollutants) to allow inter-pollutant trading (trading “dis-similar” pollutants) is growing. We examine the optimal choices of inter- and intra-pollutant trade ratios when some firms generate multiple pollutants affecting different media and when the ratios are set independent of permit caps, which are exogenously-specified by prior regulations. We find that optimal intra-pollutant trade ratios generally fail to satisfy the equi-marginal principle because there are insufficient policy instruments under exogenous caps (the Tinbergen rule is violated). Inter-pollutant trading enhances efficiency by offering an additional policy instrument: the inter-pollutant trade ratio. We also find realistic cases where inter-pollutant trading may universally improves environmental quality, in contrast to concerns that trading across environmental media may improve one medium at the expense of another. An example of nitrogen trading in the Susquehanna River Basin in Pennsylvania illustrates our results.

Keywords equi-marginal principle; instrument design; multi-pollutant markets; pollution permit markets; second-best

1. Introduction
Pollution permit trading offers the potential to improve the efficiency of pollution control by reallocating abatement effort towards more efficient abaters. Markets can also be useful in expanding the scope of environmental regulation. In the U.S., for example, many nonpoint sources of water or air pollutants (e.g., agricultural sources) are generally only regulated on a
voluntary basis. Pollution offset markets are seen as a way of voluntarily bringing these sources under the umbrella of regulation. Emissions markets have been successfully applied to manage various air pollutants (Burtraw et al. 2005), and efforts to apply markets to water are increasing (Fisher-Vanden and Olmstead 2013). While programs have historically focused on individual pollutants, many problems involve multiple, linked pollutants (US EPA 2015, 2011; NRC 2004). For example, coal-fired power plants typically release SO$_2$, NO$_X$, and CO$_2$ as joint outputs from energy production (Agee et al. 2014). These pollutants generate myriad damages across multiple environmental media, including acidification of surface waters (SO$_2$), local air quality degradation (SO$_2$ and NO$_X$), and global warming (NO$_X$ and CO$_2$). Likewise, agricultural nitrogen fertilizer use contributes simultaneously to global warming via N$_2$O emissions and to water quality impairment via nutrient loadings to waterways (Galloway et al. 2003).

The presence of pollutant linkages implies potential benefits from developing multi-pollutant strategies that take a more comprehensive approach to management (NRC 2004; Lutter and Burtraw 2002), including multi-media efforts to protect both air and water (Gray and Shadbegian 2015; Aillery et al. 2005). Indeed, multi-pollutant management is considered to be a key facet of next-generation pollution control efforts, with market-based approaches offering an important mechanism for implementation (US EPA 2011; NRC 2004). For instance, the U.S.

---

1 In the case of water pollution, section 319 of the Clean Water Act delegates the regulation of nonpoint sources to states, who have largely implemented control programs on a voluntary basis. Moreover, for both air and water pollutants, there are technical difficulties associated with nonpoint regulations due to the diffuse and largely unobservable nature of these emissions, and there is little political will for regulating agricultural sources (Shortle et al. 2012; Ribaudo 2009; Ribaudo et al. 1999).

2 In the U.S., for example, atmospheric emissions of NO$_X$ and SO$_2$ have been regulated via separate pollution trading markets under the Acid Rain Program and Cross-State Air Pollution Rule. Greenhouse gas emissions are governed in some regions by carbon trading programs, including California’s Cap and Trade Program (GWSA 2006) and the Regional Greenhouse Gas Initiative in nine northeastern states. In the case of water markets, the focus has largely been on point-nonpoint nutrient trading whereby high-cost point source abatement is exchanged for low-cost nonpoint source abatement.
Cross-State Air Pollution Rule is a multi-pollutant approach, albeit based on distinct markets for each pollutant. The European Union Emissions Trading System (EU ETS) allows trades among CO₂, nitrous oxide (N₂O), and perfluorocarbons (PFCs) by converting all pollutants into CO₂ equivalents (EC 2013). There have been calls for expanding market-based approaches that traditionally involve intra-pollutant trading, or trading “like” pollutants, to allow inter-pollutant trading, or trading across imperfectly substitutable pollutants that cannot easily be converted into equivalent units (e.g., due to heterogeneous and nonlinear environmental impacts) (NRC 2004; US EPA 2011). However, realizing the gains from multi-pollutant management requires careful policy design.

There is no reason to consider inter-pollutant trading in settings where efficiency is attainable; distinct markets for each pollutant can be constructed to be efficient or first-best, provided regulators can coordinate on the efficient permit caps and trading ratios (i.e., the rate at which abatement requirements can be traded among different sources) for the two markets. It is an open question whether gains may arise from allowing both intra- and inter-pollutant trading in cases where efficiency is unattainable, for instance due to existing regulations or other institutional constraints. Such settings are relevant because permit caps are typically not chosen based on cost-benefit analysis and therefore are unlikely to reflect the efficient level of emissions (Tietenberg 2005). Moreover, policy tools for managing pollutants are often designed using a piecemeal approach (Yaffee 1997; Lutter and Burtraw 2002). Trading ratios and other trade rules may be established after—rather than jointly with—pollution regulations, as is the case for the offset programs used to address nonpoint source water pollution and carbon emissions (Woodward 2011). These programs allow previously-regulated point sources to purchase offsets from nonpoint sources to improve the cost-effectiveness of water quality management (Wainger
and Shortle 2013; Ribaudo and Nickerson 2009; Fisher-Vanden and Olmstead 2013) or greenhouse gas emissions (GWSA 2006).

We examine the optimal design of both inter- and intra-pollutant trades when some polluters generate multiple pollutants, with a particular focus on cases where permit caps are inefficient. The impact of inefficient caps on the design of both inter- and intra-pollutant trading ratios has yet to be addressed. Lutter and Burtraw (2002) examine inter-pollutant trading using an *ad hoc* trade ratio. Others explore separate pollutant markets, with no inter-pollutant trading, when there is jointness in the production of multiple pollutants (e.g., Woodward 2011; Stranlund and Son 2015). These studies investigate the choice of emissions caps when intra-pollutant trading ratios are set at a fixed one-to-one rate due to uniform mixing of emissions for each class of pollutants. Woodward (2011) also considers whether it is more efficient to allow previously unregulated firms to sell abatement of different pollutants in only one or several distinct or separate pollution offset markets.

In the context of a single-pollutant problem, Horan and Shortle (2005) examine the design of second-best intra-pollutant trading ratios when permit caps are inefficient. They show that the trading ratio only has to manage relative emissions prices in the first-best case, whereas in the second-best case it has to manage both relative prices and effective permit endowments (i.e., the number of credits a polluter can generate through trades, which depends on the inefficient caps and the trade ratio). The second-best ratio is adjusted to address these dual tasks, but it can do neither efficiently and so market performance declines.

We extend Horan and Shortle’s (2005) analysis to the case of multiple pollutants and find inter-pollutant trading produces gains when permit caps are inefficient. These gains arise because the inter-pollutant trading ratio serves as an added tool for managing abatement allocations more
efficiently. This result is rooted in the so-called “Tinbergen Rule” that one instrument is generally needed to address each externality (Tinbergen 1956). As problems involving inefficient caps are characterized by too few instruments, adopting an additional instrument—the inter-pollutant trade ratio, which integrates markets by facilitating inter-pollutant trading—enhances efficiency relative to distinct markets that do not involve inter-pollutant trading.

We also find that the equi-marginal principle breaks down—so that effective marginal abatement costs are not equated across polluters—in both integrated and distinct markets that are designed optimally except for inefficient permit caps. This result stems from the types of second-best trade ratio adjustments that Horan and Shortle (2005) identified but did not explore analytically or numerically. Here, we find additional intra-pollutant trade ratio adjustments are required—for either distinct or integrated markets involving inefficient caps—to account for the joint production of emissions. The additional adjustments arising in the multi-pollutant case can create greater violations of the equi-marginal principle. The notion that markets should be designed to violate the equi-marginal principle is in contrast to textbook markets that promote efficiency by replicating the equi-marginal principle, and it is in contrast to prior work on inter-pollutant trading that simply imposes the equi-marginal principle, e.g., by applying one-to-one trade ratios for uniformly mixed emissions. Our numerical analysis shows there may be significant inefficiencies to imposing this principle when permit caps are inefficient.

We develop a model of multi-pollutant abatement in the next section. Section 3 derives conditions for efficient abatement, and Section 4 explores the outcome of several pollution permit trading scenarios. We highlight our analytical results using a numerical example of nitrogen trading in the Susquehanna River Basin (SRB) in Section 5. Section 6 concludes.
2. A Model of Multi-Pollutant Abatement

Consider a pollution problem whereby two environmental media (air and water) are harmed by two pollutants. For expository purposes, we focus on two different forms of nitrogen pollution: nitrous oxide emissions ($N_2O$) that contribute to atmospheric pollution, and total nitrogen loadings, defined as the quantity of nitrogen delivered to a particular water body. This pollution is produced in three sectors: an industrial sector (indexed by $I$) that produces point source emissions, denoted $e_I$; a wastewater treatment sector (indexed by $W$) that produces point source loadings, $r_W$; and an agricultural sector (indexed by $A$) that contributes nonpoint pollution to both media: air emissions, $e_A$, and water loadings, $r_A$.

Our definitions of emissions and loadings account for the effect of spatial heterogeneity on local air and water quality, which means we can consider emissions and loadings from the various sources to be uniformly mixed within an environmental media and therefore perfect substitutes in creating environmental damages. We also assume emissions and loadings from each source are deterministic. These assumptions allow us to focus on aggregate decisions at the sector level rather than focusing on individual firms within each sector. More importantly, these assumptions allow us to illustrate how the multi-pollutant case differs from textbook models that focus on a single pollutant involving deterministic and uniformly-mixed emissions.\(^3\)

Moving forward, it will be simpler to work with abatement rather than pollution levels. Define the industrial sector’s abatement as $a_{el} = e_{I0} - e_I$, where the subscript “0” denotes initial

---

\(^3\) We concentrate on optimal market design under somewhat pristine conditions (apart from exogenous permit caps) and ignore factors (e.g., uncertainty) discussed in prior work that complicate the analysis but do not help illustrate the present findings. Making the more realistic assumption that agricultural emissions are stochastic makes it more difficult to parse out the effects of multiple pollutants. This is because permit markets in such settings (based on trades of estimated or mean agricultural emissions) can only be second-best when damages are nonlinear, and generally involve a number of complex design elements (e.g., uncertainty trading ratios to adjust for risk) that significantly alter these markets relative to textbook markets (Shortle and Horan 2001). Our focus on the effects of multiple pollutants when emissions are deterministic offers a clearer comparison to textbook markets. The potential impact of stochastic agricultural pollution is described in the Discussion section.
emissions prior to abatement. The industrial abatement cost function is $C_I(a_{eI})$, where $C_I(0) = 0$ and $C'_I, C''_I > 0$. Likewise, wastewater treatment sector abatement is $a_{rW} = r_{W0} - r_W$, with the increasing, convex abatement cost function $C_{rW}(a_{rW})$, where $C_{rW}(0) = 0$. Finally, agricultural emissions and loadings abatement are $a_{eA} = e_{A0} - e_A$ and $a_{rA} = r_{A0} - r_A$, respectively, with the increasing, convex abatement cost function $C_A(a_{eA}, a_{rA})$, where $C_A(0,0) = 0$.\(^4\)

Finally, abatement reduces economic damage that depends on the ambient concentration of each pollutant. Let $E_e = E_{e0} - a_{eI} - a_{eA}$ and $E_r = E_{r0} - a_{rW} - a_{rA}$ be ambient air and water pollution, respectively, where $E_{e0} = e_{I0} + e_{A0}$ and $E_r = r_{W0} + r_{A0}$ are ambient pollution prior to abatement. Let $D_s(E_s)$ be economic damages from pollutant $s \in \{e, r\}$, with $D'_s, D''_s > 0$.

Emissions abatement benefits are then avoided economic damage, denoted $B_s(a_{eI} + a_{eA}) = D_s(E_{e0}) - D_s(E_e)$, with $B'_e(\cdot) > 0$ and $B''_e(\cdot) < 0$; loadings abatement benefits, $B_s(a_{rW} + a_{rA})$, are analogous.

Under the framework presented here, the only potential linkage between the air and water pollution problems arises through agricultural abatement costs. Assume agricultural abatement costs are not linearly separable, i.e., $\partial^2 C_A(\partial a_{eA} \partial a_{rA}) \neq 0$, to ensure the problems are linked.

Otherwise, the pollution management problem could be treated as two independent problems—one for each environmental medium.\(^5\)

---

\(^4\) Abatement costs are defined as the minimum reduction in profits required to attain a particular abatement level. Let $\pi_s(z_i)$ be profits for sector $i \in \{I, W, A\}$, $g_s(z_i)$ be the sector’s emissions or loadings of pollutant $s \in \{e, r\}$, and $z_i$ be a vector of the sector’s production and pollution control choices. Absent abatement activity, sector $i$’s profit-maximizing choice vector is $z_{i0}$. The abatement cost function for point source sector $i \in \{I, W\}$, $C_i(a_{rA})$, is obtained by choosing $z_i$ to minimize $\pi_s(z_{i0}) - \pi_s(z_i)$ subject to $a_{rA} \leq g_{rA}(z_{i0}) - g_{rA}(z_i)$. Agriculture’s abatement cost function $C_A(a_{eA}, a_{rA})$ is obtained by choosing $z_i$ to minimize $\pi_s(z_{i0}) - \pi_s(z_i)$ subject to $a_{eA} \leq g_{eA}(z_{i0}) - g_{eA}(z_i)$ and $a_{rA} \leq g_{rA}(z_{i0}) - g_{rA}(z_i)$.\(^5\) Air and water pollution linkages could also arise through the damage function $D(\cdot)$ if, for example, a fraction of all air emissions are deposited into the water resource rather than contributing only to $E_e$. This type of behavior is especially common among different types of nitrogen pollution (e.g., NO$_3$ or NOX; Galloway et al. 2003). We do not analyze such linkages here, but note that doing so requires only a simple extension of our model. Using the example of atmospheric nitrogen deposition, we could rewrite aggregate loadings as $E_s(a_{rW}, a_{eI}, a_{eA}) = r_{W0} - a_{rW} + r_{A0} - a_{rA} + \eta [e_{I0} - e_{I} + e_{A0} - e_{A}]$, where $\eta$ is the proportion of atmospheric N emissions deposited to aquatic systems. The benefits from emissions abatement would then be $B_s(a_{rW}, a_{eI}, a_{eA}) = D_s(E_{e0}) - D_s(E_e - E_s(a_{rW}, a_{eI}, a_{eA}, a_{rA})), D_s(E_{e0})$. The only difference with our current model would be an additional marginal benefit term in equations (2) and (4).
3. First-Best Control

We first characterize the efficient, or first-best, allocation of pollution control effort as a benchmark for comparison with market outcomes. The efficient outcome is defined as an allocation of pollution control effort that maximizes social net benefits

\[
\max_{a_{el}, a_{eW}, a_{eA}, a_{rA}} V = B_e(a_{el} + a_{eA}) + B_r(a_{eW} + a_{rA}) - C_I(a_{el}) - C_W(a_{eW}) - C_A(a_{eA}, a_{rA}).
\]

Assuming an interior solution, the first-order conditions (FOCs) for problem (1) are

\[
\frac{\partial V}{\partial a_{el}} = 0 \Rightarrow C'_e = B'_e,
\]

\[
\frac{\partial V}{\partial a_{eW}} = 0 \Rightarrow C'_W = B'_r,
\]

\[
\frac{\partial V}{\partial a_{eA}} = 0 \Rightarrow \frac{\partial C_A}{\partial a_{eA}} = B'_e,
\]

\[
\frac{\partial V}{\partial a_{rA}} = 0 \Rightarrow \frac{\partial C_A}{\partial a_{rA}} = B'_r.
\]

The FOCs (2)–(5) state the familiar result that, at the first-best abatement levels \(a^{*}_{el}, a^{*}_{eW}, a^{*}_{eA}, a^{*}_{rA}\) and \(a^{*}_{rA}\), each sector’s marginal abatement costs equal the marginal benefits from abatement.

Additional insight is obtained by manipulating (2)–(5) to yield the following modified equi-marginal condition,

\[
\frac{C'_e}{B'_e} = \frac{\partial C_A}{\partial a_{eA}} = \frac{C'_W}{B'_r} = \frac{\partial C_A}{\partial a_{rA}} = 1,
\]

so that the effective marginal cost of abatement—measured by the marginal cost normalized by the marginal avoided damages from abatement—is equalized across all sources under the efficient outcome. The first and third equalities in (6) imply the conventional equi-marginal condition: marginal abatement costs should be equated within each environmental medium. The
second equality extends the equi-marginal principle across media by normalizing costs in a way that treats abatement benefits in the different media as fungible or perfectly substitutable. The final equality indicates marginal costs equal marginal benefits in each case. While not a surprising result, this modification contrasts with current regulatory approaches (e.g., distinct markets for different pollutants) that treat abatement of distinct pollutants as non-substitutable even when those pollutants arise from the same source.

4. Market Trading Scenarios

We now consider the outcome under various trading scenarios that differ along two dimensions. First, pollution trading can occur either in distinct markets for each pollutant that allow only intra-pollutant trading (reflecting current market-based approaches) or in an integrated, multi-pollutant market that allows both intra- and inter-pollutant trading. Second, point source pollutant caps can either be chosen optimally in conjunction with trading ratios, or they can be set exogenously relative to trading ratios. We begin by defining the market responses in a general model of trading, as these responses will be used to construct the various trading scenarios.

4.1. Market Responses in a General Model of Pollution Trading

We adopt a general model of trading by defining sector-specific permits for each pollutant. Agricultural permits are denoted \( \hat{e}_A \) and \( \hat{r}_A \), with initial allocations \( \hat{e}_{A0} \) and \( \hat{r}_{A0} \) and permit prices \( p_{eA} \) and \( p_{rA} \). The agricultural sector is not initially regulated, i.e., \( \hat{e}_{A0} = e_{A0} \) and \( \hat{r}_{A0} = r_{A0} \), and so farmers have initial rights to pollute. Point source permits are denoted \( \hat{e}_I \) and \( \hat{r}_W \), with associated initial allocations or permit caps \( \hat{e}_{I0} \) and \( \hat{r}_{W0} \) and permit prices \( p_{eI} \) and \( p_{rW} \). Point sources initially face binding regulations, i.e., \( \hat{e}_{I0} < e_{I0} \) and \( \hat{r}_{W0} < r_{W0} \) but they may purchase permits (or offsets) from other sources to pollute more and reduce abatement costs.
We examine two types of market structures. First is an integrated market that allows both intra- and inter-pollutant trading. Industrial emissions are defined as the numeraire pollutant, and trades are guided by trading ratios that define the number of permits that must be purchased for an industrial source to increase emissions by one unit. Three ratios are required: an intra-pollutant ratio for emissions, \( \tau_{eA,eI} = \frac{d\hat{e}_A}{d\hat{e}_I} \), and two inter-pollutant ratios, \( \tau_{rA,el} = \frac{|d\hat{r}_A/d\hat{e}_I|} {\tau_{rA,el}} \) and \( \tau_{rW,el} = \frac{|d\hat{r}_W/d\hat{e}_I|} {\tau_{rW,el}} \). These ratios can be used to define the trade ratios for the other potential trades: an intra-pollutant ratio for loadings, \( \tau_{rA,rW} = \frac{|d\hat{r}_A/d\hat{r}_W|}{\tau_{rA,el}} \) and two remaining inter-pollutant ratios, \( \tau_{rA,eA} = \frac{|d\hat{r}_A/d\hat{e}_A|}{\tau_{rA,el}} \) and \( \tau_{eA,rW} = \frac{|d\hat{e}_A/d\hat{r}_W|}{\tau_{rA,el}} \). The market clearing condition for this case is

\[
\begin{align*}
\hat{e}_{10} + \frac{\hat{e}_{40}}{\tau_{eA,el}} + \frac{\hat{r}_{W0}}{\tau_{rW,el}} + \frac{\hat{r}_{40}}{\tau_{rA,el}} &= \left( e_{10} - a_{el} \right) + \left( e_{40} - a_{el} \right) / \tau_{eA,el} \\
&\quad + \left( r_{W0} - a_{el} \right) / \tau_{rW,el} + \left( r_{40} - a_{el} \right) / \tau_{rA,el}
\end{align*}
\]

where the left hand side (LHS), denoted \( Q_e \), is the effective aggregate permit cap denominated in terms of industrial emissions. Notice the effective cap or endowment depends on the trade ratios.

Now consider the case of distinct markets, defined as an emissions market and a loadings market in which no inter-pollutant trades are allowed—only intra-pollutant trading occurs within each market. Point source permits serve as a numeraire in their respective markets, with one intra-pollutant trade ratio required for each market: \( \tau_{eA,el} = \frac{d\hat{e}_A/d\hat{e}_I} {\tau_{eA,el}} \) and \( \tau_{rA,rW} = \frac{|d\hat{r}_A/d\hat{r}_W|} {\tau_{rA,el}} \). There is also a distinct market-clearing condition for each market:

\[
\begin{align*}
\hat{e}_{10} + \frac{\hat{e}_{40}}{\tau_{eA,el}} &= \left( e_{10} - a_{el} \right) + \left( e_{40} - a_{el} \right) / \tau_{eA,el} \\
\hat{r}_{W0} + \frac{\hat{r}_{40}}{\tau_{rA,el}} &= \left( r_{W0} - a_{el} \right) + \left( r_{40} - a_{el} \right) / \tau_{rA,el}
\end{align*}
\]

where \( Q_e \) and \( Q_r \) represent the effective aggregate permit caps for each market.

Now consider each sector’s decisions. Each sector chooses abatement to minimize abatement costs plus the cost of purchasing permits. We show in the Appendix that these
problems can be written as follows, regardless of whether the markets are integrated or distinct:

\[
\min_{a_{el}} C_I(a_{el}) + p_{el} [e_{I0} - a_{el} - \hat{e}_{I0}]
\]

\[
\min_{a_{rw}} C_W(a_{rw}) + p_{rw} [r_{W0} - a_{rw} - \hat{r}_{W0}]
\]

\[
\min_{a_{el},a_{ra}} C_A(a_{el},a_{ra}) + p_{el} [e_{A0} - a_{el} - \hat{e}_{A0}] + p_{ra} [r_{A0} - a_{ra} - \hat{r}_{A0}]
\]

with the following FOCs for interior solutions

\[
C'_I = p_{el}
\]

\[
C'_W = p_{rw}
\]

\[
\frac{\partial C_A}{\partial a_{el}} = p_{el}
\]

\[
\frac{\partial C_A}{\partial a_{ra}} = p_{ra}.
\]

Conditions (11)–(14) simply state that, at the optimum, each sector’s marginal abatement cost equals the permit price of the abated pollutant. Further, we show in the Appendix that additional market equilibrium conditions relate the trade ratios to permit price ratios, consistent with prior work on markets for a single pollutant (e.g., Malik et al. 1993). Using these relations from the Appendix, along with (11)–(14), the market equilibrium conditions for an integrated market are:

\[
\tau_{el,el} = \frac{p_{el}}{p_{el}} = \frac{C'_I}{\frac{\partial C_A}{\partial a_{el}}}, \quad \tau_{ra,el} = \frac{p_{el}}{p_{ra}} = \frac{C'_I}{\frac{\partial C_A}{\partial a_{ra}}}, \quad \tau_{rw,el} = \frac{p_{el}}{p_{rw}} = \frac{C'_I}{C'_W}.
\]

The market equilibrium conditions for distinct markets are

\[
\tau_{el,el} = \frac{p_{el}}{p_{el}} = \frac{C'_I}{\frac{\partial C_A}{\partial a_{el}}}, \quad \tau_{ra,el} = \frac{p_{el}}{p_{ra}} = \frac{C'_I}{\frac{\partial C_A}{\partial a_{ra}}}, \quad \tau_{rw,el} = \frac{p_{el}}{p_{rw}} = \frac{C'_I}{C'_W}.
\]

The abatement choices that solve the relevant market equilibrium conditions (15) and the relevant market-clearing condition (7) can be expressed as responses to the policy variables,
\[ a_{el}(\tau, \hat{e}), a_{eA}(\tau, \hat{e}), a_{rA}(\tau, \hat{e}), \text{ and } a_{rW}(\tau, \hat{e}), \] where \( \tau = [\tau_{rA,el} \ \tau_{rW,el} \ \tau_{eA,el}] \) in the integrated market scenario and \( \tau = [\tau_{eA,el} \ \tau_{rA,rW}] \) in the distinct markets scenario, whereas \( \hat{e} = [\hat{e}_0 \ \hat{r}_W] \) in both scenarios. Note that each sector’s behavior depends on the policy variables associated with both air and water pollution. It is obvious that this should be the case for an integrated market, but it is also true for distinct markets due to the agricultural source participating in both markets; its non-separable abatement costs imply its decisions are dependent across markets. It is analytically intractable to identify the signs of the comparative statics results for the abatement response functions (which stem from applying Cramer’s Rule to a 4 \( \times \) 4 matrix). However, prior work on trading a single pollutant across two sectors indicates that a larger ratio \( \tau_{y,z} \) generally makes it more expensive for sector \( z \) to trade to reduce abatement (Horan and Shortle 2015).


We begin our analysis of optimal permit market design with the case where permit caps are endogenously chosen. We start with this case because it most directly relates to prior work on market design (e.g., Montero 2001) and because prior work has not formally examined the optimal choices of intra-pollutant trade ratios in a multi-pollutant setting. The planner’s problem, after substituting the behavioral responses into \( V \), is

\[
\max_{\tau, \hat{e}} V = B_e(a_{el}(\tau, \hat{e}) + a_{eA}(\tau, \hat{e})) + B_r(a_{rW}(\tau, \hat{e}) + a_{rA}(\tau, \hat{e})) - C_I(a_{el}(\tau, \hat{e})) - C_W(a_{rW}(\tau, \hat{e})) - C_A(a_{eA}(\tau, \hat{e}), a_{rA}(\tau, \hat{e}))
\]

Problem (16) is written generally to reflect both market scenarios. The FOC for any relevant policy parameter \( u \), defined as a scalar element of the policy vectors, is

\[
\frac{\partial V}{\partial u} = \left[ B_e' - C_I' \right] \frac{\partial a_{el}}{\partial u} + \left[ B_r' - C_W' \right] \frac{\partial a_{rW}}{\partial u} + \left[ B_r' - C_A' \right] \frac{\partial a_{rA}}{\partial u} = 0 \ \forall u.
\]

Comparing (17) with the efficiency conditions (2)–(5), it is clear that an efficient market design
causes each of the four bracketed terms in (17) to vanish. Ensuring such an outcome generally requires four instruments—one instrument for each externality, consistent with the Tinbergen rule (Tinbergen 1956). Note that four instruments are available in both the integrated and distinct market scenarios.6

For the integrated market scenario, FOC (17) can be written in terms of three trading ratios by using the sectors’ FOCs along with the market equilibrium relations (15a):

\[
(18a) \left[ B'_e - \frac{P_{el}}{\tau_{el,el}} \right] \frac{\partial a_{el}}{\partial u} + \left[ B'_r - \frac{P_{el}}{\tau_{el,el}} \right] \frac{\partial a_{el}}{\partial u} + \left[ B'_r - \frac{P_{el}}{\tau_{el,el}} \right] \frac{\partial a_{el}}{\partial u} + \left[ B'_r - \frac{P_{el}}{\tau_{el,el}} \right] \frac{\partial a_{el}}{\partial u} = 0 \quad \forall u.
\]

The bracketed terms in (18a), and hence in (17), will vanish to yield an efficient outcome when \( \tau_{el,el} = \frac{B'_e}{B'_e} \) and \( \tau_{el,el} = \frac{B'_e}{B'_e} = 1 \) and \( \hat{e} \) is set such that \( p_{el} = B'_e \) (see footnote 6). Using the market equilibrium conditions, we can use the trading ratio results to derive the remaining equilibrium prices \( \tau_{el,el} = \frac{B'_e}{B'_e} \) and \( \tau_{el,el} = \frac{B'_e}{B'_e} = 1 \), as well as equivalent inter-pollutant trading ratios, \( \tau^* = \tau_{el,el} = \tau_{el,el} = \tau_{el,el} \) for one-to-one intra-pollutant trading for emissions, \( \tau^* = \frac{p_{el}}{p_{el}} = 1 \), and loadings, \( \tau_{el,el} = \frac{B'_e}{B'_e} \) for one-to-one intra-pollutant trading is consistent with current approaches for uniformly-mixed pollutants. The optimal inter-pollutant trade ratio, based on relative economic benefits, contrasts with the standard approach in current multi-pollutant markets of setting the ratio.

---

6 As depicted in problem (16), the integrated market scenario involves five policy choices to manage four abatement levels. In this case it is equivalent to transform the problem slightly by treating \( Q \) from the market clearing condition (7a) as the policy variable rather than individual permit levels. Then the abatement response functions would take the form \( a_{el} (\tau, Q), a_{el} (\tau, Q), a_{el} (\tau, Q), \) and \( a_{el} (\tau, Q), \) and the choice of \( Q \) along with the three trading ratios would be sufficient to manage the four abatement levels. Optimizing (16) with respect to \( Q \) and \( \tau \) yields the first-best levels \( Q^* \) and \( \tau^* \). Then the initial permit allocation \( \hat{e}_{I0} \) and \( \hat{e}_{W0} \) may be set at any combination to satisfy \( Q^* = \hat{e}_{I0} + \hat{e}_{I0}/\tau_{el,el} + \hat{e}_{el,el}/\tau_{el,el} + \hat{e}_{el,el}/\tau_{el,el} \).
according to the pollutants’ relative physical or chemical qualities.\footnote{For example, different types of GHGs are traded based on their global warming potential, denominated in units of “carbon dioxide equivalents.” Trades governed by these types of trade ratios are unlikely to be cost-effective as they ignore the economic characteristics of pollution that vary across pollutants (Schmalensee 1993; Muller 2012).} However, the efficient ratio is consistent with prior economic research that says efficient trades should occur at the marginal rate of substitution of damages (e.g., Schmalensee 1993; Lutter and Burtraw 2002).

For the distinct markets scenario, FOC (17) can be written in terms of two trading ratios by using the sectors’ FOCs along with the market equilibrium relations (15b):

\begin{equation}
\left[ B'_e - P_{el} \right] \frac{\partial a_{el}}{\partial u} + \left[ B'_e - P_{el} \right] \frac{\partial a_{el}}{\partial u} + \left[ B'_r - P_{rW} \right] \frac{\partial a_{rW}}{\partial u} + \left[ B'_r - P_{rW} \right] \frac{\partial a_{rW}}{\partial u} = 0 \forall u
\end{equation}

The bracketed terms in (18b), and hence in (17), will vanish to yield an efficient outcome when

\[ \frac{\tau_{eA,el}}{B'_e} = 1, \quad \frac{\tau_{rA,rW}}{B'_r} = 1, \quad \hat{e}_{10}^* \text{ is set such that } P_{el} = B'_e, \text{ and } \hat{r}_{0W}^* \text{ is set such that } P_{rW} = B'_r. \]

Using the market equilibrium conditions, we can use the trading ratio results to derive the remaining equilibrium prices

\[ p_{el} = \frac{P_{el}}{\tau_{eA,el}} = B'_e \quad \text{and} \quad p_{rA} = \frac{P_{rW}}{\tau_{rA,rW}} = B'_r. \]

This solution ensures the point-nonpoint emissions and loadings trading ratios are

\[ \tau_{eA,el}^* = \tau_{rA,rW}^* = 1. \]

There is one important caveat to these results for both the integrated market and the distinct markets. Let \( E_{e}^* = E_{e0} - a_{el}^* - a_{el}^* \) and \( E_{r}^* = E_{r0} - a_{rW}^* - a_{rW}^* \) be the efficient levels of total emissions and loadings, respectively. Assuming agricultural sources are not initially regulated so that they have implicit initial permit caps of \( e_{A0} \) and \( r_{A0} \), then the efficient permit caps for the emissions and loadings by the industrial and wastewater treatment sectors, respectively, are

\[ \hat{e}_{10}^* = E_{e0} - e_{A0} \quad \text{and} \quad \hat{r}_{0W}^* = E_{r0} - r_{A0}. \]

Note that \( \hat{e}_{10}^* \leq E_{e}^* \) and \( \hat{r}_{0W}^* \leq E_{r}^* \) are required to obtain the efficient outcome with \( \hat{e}_{A0} = e_{A0} \) and \( \hat{r}_{A0} = r_{A0} \); in other words, initial nonpoint source emissions and loadings must not exceed efficient levels. Otherwise, initial nonpoint source emissions
and/or loadings are so large that the first-best outcome cannot be attained simply by regulating point source emissions and loadings. In such instances, agriculture must be regulated, $\dot{e}_{A0} < e_{A0}$ and/or $\dot{r}_{A0} < r_{A0}$, to obtain the first-best outcome. Otherwise, the permit levels are effectively constrained as in the exogenous permit cap scenarios described in section 4.3 below.

The two market scenarios imply the following result:

**Result 1.** Suppose point source permit caps are endogenously chosen while nonpoint source caps are essentially set at unregulated levels. Either integrated or distinct markets can be efficient (provided unregulated agricultural pollution is not too great) because the tradeable permit markets can replicate the first-best equi-marginal principle. Moreover, one-to-one trading is optimal for intra-pollutant trades involving uniformly-mixed pollutants, consistent with the traditional equi-marginal principle.

Result 1 indicates we could obtain an efficient outcome by integrating permit markets, but there is no need for this integration provided point source permit caps are chosen optimally to reflect the linkages created by the agricultural sector (or, more generally, any sector that pollutes in multiple markets). Moreover, neither the linkages nor decisions on market integration affect the standard result of one-to-one intra-pollutant trades for uniformly-mixed pollutants.

In practice, market caps are not typically chosen efficiently, but instead are set outside the market to meet environmental or human health standards. We now turn to the more realistic case of exogenously-defined pollution caps to examine trading in a second-best setting.

### 4.3. Market Design with Exogenous Permit Caps

Consider the optimal choice of $\tau$ given that $\dot{e}$ has already been exogenously specified, likely at a
sub-optimal value. The objective function for this new problem, in which we simplify the 
notation by suppressing $\hat{e}$, is

$$\begin{align*}
\max V = B_e (\tau_{cd}(\tau) + a_{cd}(\tau)) + B_r (\tau_{W}(\tau) + a_{cd}(\tau)) \\
- C_1 (a_{cd}(\tau)) - C_W (\tau_{W}(\tau)) - C_A (a_{cd}(\tau), a_{cd}(\tau))
\end{align*}$$

(19)

The first order conditions are analytically equivalent to (17), but now $u$ is only defined as a scalar 
element of the vector $\tau$. This means there are no longer four policy instruments—in either type of 
market—to ensure that each of the bracketed terms in (17) vanishes to produce the efficient 
outcome; the Tinbergen rule is violated and the resulting market solutions cannot be first-best.

First consider the case of integrated markets. We show in the Appendix that the optimal 
trading ratios are defined implicitly by the following relations:

$$\begin{align*}
\tau_{rW,cd} &= \frac{B'_e}{C'_W} + \frac{B'_e - C'_W}{C'_W} \rho_{rW,cd}(\tau) + \frac{B'_e}{C'_W} \rho_{rW,cd}(\tau) - 1 + \frac{B'_r}{C'_W} \rho_{rW,cd}(\tau) + \frac{B'_r}{C'_W} \rho_{rW,cd}(\tau)
\end{align*}$$

(20a)

$$\begin{align*}
\tau_{eA,cd} &= \frac{B'_e}{C'_A} + \frac{B'_e - \partial C_A / \partial a_{cd}}{C'_A} (\tau_{A,cd} + \tau_{A,cd}) + \frac{B'_e}{C'_A} (\rho_{eA,cd}(\tau) - 1) + \frac{B'_e}{C'_A} \rho_{eA,cd}(\tau)
\end{align*}$$

(20b)

$$\begin{align*}
\tau_{rA,cd} &= \frac{B'_e}{C'_A} + \frac{B'_e - \partial C_A / \partial a_{cd}}{C'_A} (\tau_{A,cd}) + \frac{B'_e}{C'_A} (\rho_{rA,cd}(\tau) - 1) + \frac{B'_e}{C'_A} \rho_{rA,cd}(\tau)
\end{align*}$$

(20c)

where $\rho_{j,y,z}(\tau) = \sum_k (\partial a_j / \partial \tau_{y,z}) / \tau_{j,y,z}$ (for $j, y, z \in \{eI, eA, rW, rA\}$ and $\tau_{j,y,z} = 1$) is the marginal 
abatement response of sector $j$ to trade ratio $\tau_{j,y,z}$, relative to the marginal responses in all other 
sectors (with all effects denominated in terms of sector $z$’s permits). The first RHS term in (20) is
the ratio of marginal benefits of abatement in sector $z$ relative to the marginal cost of abatement in sector $y$. This term is of the same form as the first-best trade ratio $\tau_{y,z}$ presented above, given that the first-best outcome equates sector $y$’s marginal benefits and marginal costs of abatement.

The remaining RHS terms in (20) are adjustments to address inefficiencies from having too few instruments to perfectly control each sector’s abatement. Recall from the market equilibrium condition (15) that trade ratios determine relative permit prices, and from the market clearing condition (7) that they determine effective permit endowments. Trade ratios only have to manage relative prices when permit levels are chosen optimally to manage endowments. But if permits levels are exogenous, then trade ratios must perform both tasks, in which case neither task is performed efficiently. The second RHS term in (20) reflects an inefficient deviation in the marginal benefits and marginal costs of abatement in sector $y$, arising from an inability to adjust permit endowments to equate these values. We expect $\rho_{j,y,z}(\tau)$ to be negative, which means $\tau_{y,z}$ is optimally decreased (increased) to incentivize more (less) abatement in sector $y$ whenever the inefficiencies increase (decrease) the marginal benefits of abatement in sector $y$ relative to the marginal abatement costs. This term was identified by Shortle and Horan (2005) and described as a trade ratio modification to move the trading equilibrium “close” to the first-best optimum.

Two additional terms arise here due to the multi-pollutant nature of the current problem. Specifically, the third and fourth terms in (20) reflect the economic impacts of inefficient behavioral responses to $\tau_{y,z}$ outside of sector $y$, as manifested through the $\rho_{j,y,z}(\tau)$ terms. In other words, trying to adjust the effective endowment in one sector causes an inefficient behavioral response in the other sector due to the markets being linked.8

A key result from equation (20) is that intra-pollutant trading at a one-to-one rate (i.e.,

---

8 Note that the first-best trade ratios (18) arise as a special case of (20) when permit levels are set to ensure the behavioral responses yield $\rho_{j,y,z}(\tau) = 0 \forall j \neq z,y$, such that $\rho_{y,y,z}(\tau) = 1$ (i.e., at the margin, a change in $\tau_{y,z}$ for $y \neq z$ only guides trades between sectors $y$ and $z$, with no effect on other sectors’ abatement), and $\rho_{z,y,z}(\tau) = -\tau_{y,z} \forall j \neq z,y$. 

\( \tau_{eA, eI} = 1; \tau_{rA, rW} = \tau_{rA, eI}/\tau_{rW, eI} = 1 \) is unlikely to be optimal, even though pollution is uniformly mixed within each environmental medium. More generally, standard notions of the equi-marginal principle break down. Note that this result holds even in the special case where agricultural abatement costs are linearly separable due to the fact that the integrated market links behaviors affecting both environmental media.

Now consider the case of distinct markets. We begin by noting this scenario has one less policy tool than the integrated market scenario (two ratios here versus three in the integrated market).\(^9\) This means the distinct market scenario cannot yield larger net benefits than the integrated market case by Le Châtelier’s Principle (Samuelson 1947), and so it may be more apt to describe this scenario as third best.\(^{10}\) Even so, the optimality conditions are still given by (17).

Following the same approach used above, the optimal trade ratios for this case are (20b) and

\[
(20d) \quad \tau_{rA, rW} = \frac{B'_r - \left( \frac{\partial C_A}{\partial a_{rA}} \right)_{rA,rA,rW}(\tau)}{\frac{\partial C_A}{\partial a_{rA}} - \frac{\partial C_A}{\partial a_{rW}}} + \frac{B'_r \left[ 0_{rW, rA, rW}(\tau) - 1 \right]}{\frac{\partial C_A}{\partial a_{rA}}} + \frac{B'_r \left[ 0_{eA, rA, rW}(\tau) + \rho_{eA, rA, rW}(\tau) \right]}{\frac{\partial C_A}{\partial a_{rW}}}.
\]

The interpretation of this intra-pollutant ratio is the same as above, as is the result that intra-pollutant trading at a one-to-one rate is unlikely to be optimal. Note that the third and fourth RHS terms in (20b) and (20d) do not vanish, even though there are no explicit market linkages in the present case of distinct markets. The reason is that agricultural abatement costs are not linearly separable, so that incentives in one market affect agricultural choices in the other market. If agricultural abatement costs were linearly separable, such that the markets were truly separate.

\(^9\) Recall this was also true when permit levels were endogenous, although there were no efficiency implications in that case since there were sufficient numbers of controls in each market scenario to attain a first-best outcome (see footnote 6). Making permit levels exogenous results in two fewer controls in each market scenario, imposing efficiency-reducing restrictions for each scenario.

\(^{10}\) Beavis and Walker (1983) use the terminology “third-best” to describe a pollution control strategy that optimizes economic welfare in the face of constraints on the design or implementation of the strategy that prevent the solution from even being second-best.
with \( \rho_{z,y,z}(\tau) = 1 \), \( \rho_{y,y,z}(\tau) = -\tau_{y,z} \), and \( \rho_{j,y,z}(\tau) = 0 \ \forall j \neq z,y \), then setting \( \tau_{eA,eI} = \tau_{eA,rW} = 1 \) yields the least-cost allocation associated with the chosen permit levels.

The results of this section are summarized as follows:

**Result 2.** Suppose point source permit caps and nonpoint source caps are set exogenously relative to the trading program. We find one-to-one intra-pollutant trading is sub-optimal in this case, even when pollutants within a particular medium are uniformly-mixed. More generally, the equi-marginal principle does not apply when caps are exogenous. Finally, whereas integrated markets can only be second best, distinct markets are likely to be even less efficient.

In practice, caps are set exogenously for individual pollutants, and then distinct markets are implemented with trading occurring on a one-to-one basis for uniformly mixed pollutants. Result 2 indicates that such a market design is not even third-best.

**5. Numerical Model: Multi-Pollutant Trading in the Susquehanna River Basin**

We now illustrate the theory using a model of multi-pollutant trading in the Pennsylvania portion of the SRB (Figure 1). The SRB is the Chesapeake Bay’s largest drainage basin, contributing about 60 percent of the total streamflow and nearly 46 percent of the nitrogen loads to the Bay (US EPA 2010). Most of the SRB is in Pennsylvania, which is the major source of the SRB’s nutrient inputs. Pennsylvania established a nutrient water quality trading program in 2005 under its 2004 Chesapeake Bay Tributary Strategy (Shortle 2012). Trading activity, while sparse initially, has recently increased due to stringent new caps imposed by the 2010 Chesapeake Bay Total Maximum Daily Load (TMDL). The Pennsylvania portion of the SRB also features numerous point and nonpoint sources of greenhouse gases (GHGs). Pollution in the form of
GHG emissions and water quality loadings poses a major threat to environmental quality and economic activities in the Chesapeake Bay and the surrounding airshed (Birch et al. 2011).

5.1. Model Specification and Calibration

We begin by specifying and calibrating a numerical model of abatement costs and benefits for the SRB. The emissions data used to calibrate our model is expressed in metric tons of carbon equivalents (mtCO$_2$e), and the loadings data is expressed in pounds of N. We use these original units when calibrating our model, then scale the parameters so that emissions are expressed in millions of mtCO$_2$e and loadings are expressed in thousands of metric tons of N (mtN). The scaled parameters are reported in Table 1.

Let $C_I(a_{el}) = (\psi / 3)a_{el}^3$ be the industrial sector’s abatement cost so that marginal abatement costs are $C'_I(a_{el}) = \psi a_{el}^2$. Assuming a marginal abatement cost of $35 at an abatement level of $a_{el} = 19.8$ million mtCO$_2$e (RGGI 2014; US EPA 2013), we solve for $\psi = 8.9 \times 10^4$.

Similarly, let $C_W(a_{rw}) = (\phi / 3)a_{rw}^3$ be the wastewater treatment sector’s abatement costs so that marginal abatement costs are $C'_W(a_{rw}) = \phi a_{rw}^2$. Kaufman et al. (2014) estimate the marginal abatement cost for point sources in the Pennsylvania portion of the SRB to be $15 at an abatement level $a_{rw} = 4.99 \times 10^6$ lbN. Substituting this information into the marginal abatement cost relation and solving yields $\phi = 6.42 \times 10^6$.

Finally, let $C_A(a_{et}, a_{ra}) = (\alpha / 3)a_{et}^3 + (\beta / 3)a_{ra}^3 - \gamma a_{et}a_{ra}$ be the agricultural sources’ abatement cost function. Following Woodward (2011), we assume agricultural emissions and
loadings abatement are complements, with $\gamma > 0$ representing the degree of complementarity.$^{11}$ We are unaware of any empirical measurement of the complementarity between agricultural emissions and loadings abatement, so we assume $\gamma = 1.1 \times 10^7$. This is small enough that marginal abatement costs do not become negative over reasonable ranges of abatement.$^{12}$ The parameters $\alpha$ and $\beta$ are calibrated by substituting values for marginal abatement costs and abatement levels into the marginal cost relations. For loadings, we use Kaufmann et al.’s (2014) marginal abatement cost value of $11.58$/lbN at an abatement level of 20,955,000 lbN. Using these values in the marginal abatement cost relation $\frac{\partial C_A}{\partial a_{rA}} = \beta a_{rA}^2 - \gamma a_{eA}$, we solve for $\beta = 3.7 \times 10^5$.

We use the agricultural loadings abatement value to determine the corresponding level of emissions abatement of $a_{eA} = 734,451$ mt$\text{CO}_2\text{e}$. Assuming a marginal abatement cost of $10 (Golub et al. 2009), we use the relation $\frac{\partial C_A}{\partial a_{eA}} = \alpha a_{eA}^2 - \gamma a_{rA}$ to solve for $\alpha = 2.08 \times 10^8$.

We assume the damage function for emissions takes the form $D_e(E_e) = \varepsilon E_e$ since GHGs are a globally mixed pollutant and the emissions from the Pennsylvania portion of the SRB represent a small portion of the world’s GHG emissions. We set $\varepsilon = 1.4 \times 10^7$ in accordance with Tol’s (2005) median estimate of the marginal damage value of CO$_2$ emissions.

Finally, damages from loadings take the form $D_r(E_r) = (\nu/2)E_r^2$, so that $D'_r(E_r) = \nu E_r$ where $\nu > 0$ is a parameter. Kaufman et al. (2014) estimate marginal damages for total N loadings in the Chesapeake Bay to be $3.37$ at an aggregate emissions level of $E_r = 9.9 \times 10^7$ lbN.

Substituting these values into the marginal damage relation, we solve for $\nu = 1.65 \times 10^5$.

$^{11}$ In other settings, agricultural abatement of the linked pollutants may be substitute. For example, no-till farming can reduce N loss from farm fields, but increased weed pressure may result in increased herbicide use, and hence increase herbicide loadings to waterways (Soane et al. 2012).

$^{12}$ We performed sensitivity analysis and found that our numerical results are largely insensitive to $\gamma$.

$^{13}$ Specifically, we used the initial loadings abatement value along with transport coefficients from the USGS SPARROW model (Ator et al. 2011) to calculate the average change in applied nitrogen for cropland in the SRB, assuming all abatement was due to changes in nitrogen application. We then follow the approach of Reeling and Gramig (2012) of using the DAYCENT model (NREL 2011) to estimate emissions abatement associated with this change in nitrogen applications.
5.2. Simulation Results

We simulate a variety of trading scenarios for the SRB using MATLAB (Mathworks, Inc. 2016); the results are presented in Table 2. In each scenario, the inter-pollutant trade ratio denotes trades involving equivalent units (e.g., metric tons-for-metric tons). Scenario 1, or the efficient scenario, represents a first-best or efficiently designed integrated pollution market. Although not reported in Table 2, the initial emissions and loading caps, \( \hat{e}_{I0} \) and \( \hat{r}_{W0} \), are chosen to ensure

\[
p_{el} = p_{eA} = B'^*_{e} \quad \text{and} \quad p_{rw} = p_{rA} = B'^*_{r} \quad \text{in the market outcome, given that nonpoint sources have an implicit right to pollute. A distinct market with the same intra-pollutant trade ratios can also yield the efficient outcome, provided the caps are adjusted to account for no inter-pollutant trading.}
\]

The first-best results indicate industry is the least cost source of emissions abatement, as evidenced by the large relative abatement allocation to industry, whereas agriculture is the least cost source of loadings abatement. These results are due to agriculture’s relatively high marginal cost of abating emissions, but relatively low marginal cost of abating loadings. The efficient inter-pollutant trade ratio is less than one to encourage more abatement in the loadings sector at the margin, due to convex damages from loadings (whereas marginal damages from the SRB’s emissions are constant). The efficient intra-pollutant trade ratios are unity, as is expected.

We compare the efficient outcome to the more realistic integrated and distinct market scenarios in which the initial permit caps for each pollutant have been set exogenously and sub-optimally, e.g., by different agencies regulating point sources in each sector. These sub-optimal point source permit caps represent the only differences in initial regulations relative to the efficient outcome, as we have assumed the agricultural sector is not initially regulated in either

---

14 Notice that three trade ratios are presented for each scenario: one inter-pollutant trade ratio and two intra-pollutant trade ratios. In the analytical section above, we optimized over the inter-pollutant ratio \( \tau_{el/eA} \) rather than the intra-pollutant ratio \( \tau_{el/eA} \). This is of no consequence, since any three ratios give rise to all other ratios. We focus on the intra-pollutant ratio in our numerical results to make comparisons with the case in which one-to-one intra-pollutant trade ratios are imposed exogenously.
Therefore, to facilitate comparison with the first-best case, let the point source permit caps be $\hat{e}_{t0} = \zeta_e \hat{e}_{t0}^*$ and $\hat{r}_{w0} = \zeta_r \hat{r}_{w0}^*$, where $\zeta_s$ (for $s \in \{e, r\}$) is a parameter indicating the degree of regulation relative to the efficient case. A value of $\zeta_s = 1$ represents an efficient cap on point sources of pollutant $s$, whereas $\zeta_s < 1$ ($> 1$) implies an inefficiently strict (lax) cap. Note that an efficient cap in only one sector (e.g., $\zeta_e = 1$, $\zeta_r \neq 1$ or vice versa) will not yield an efficient allocation of either type of pollutant since agricultural choices respond to incentives in both markets. Efficiency is only obtained when $\zeta_e = \zeta_r = 1$. We set $\zeta_e = \zeta_r = \zeta$ in Table 2 because our interest is in highlighting how traditional results change when caps are set exogenously, rather than in exploring every qualitative combination of $\zeta_e$ and $\zeta_r$ relative to unity.

Table 2 illustrates results for $\zeta = 0.95$ (a strict cap) and $\zeta = 1.05$ (a lax cap), with three scenarios examined for each value of $\zeta$. The first scenario involves no trading, whereby point sources must satisfy the caps on their own while nonpoint sources remain unregulated. The no-trading scenarios are insightful because comparing these to the various trading scenarios illustrates the economic and ecological performance of trading. The second scenario considered involves optimally chosen trade ratios, whereby all trading ratios are optimally chosen to satisfy the cap given the type of market being considered (integrated versus distinct). The third scenario considered involves one-to-one intra-pollutant ratios, whereby all intra-pollutant ratios are exogenously set equal to one. Both integrated and distinct markets are considered for each of the trading scenarios.

First consider the no-trading scenarios 2 and 7. Social net benefits are negative for both $\zeta = 0.95$ and $\zeta = 1.05$ because the cost of point sources satisfying the abatement requirements on their own greatly exceeds the social benefits. The costs are particularly large under the case of strict caps, with industry abating 61 percent more and wastewater treatment plants abating 649
percent more than in the efficient outcome. This means abatement costs increase significantly for all point sources, with wastewater treatment plants bearing the greatest costs of over-regulation absent trading.

The environmental performance of the no-trading scenarios is measured by the aggregate pollutant levels arising in these scenarios, which equals the sum of the point source emissions cap and the nonpoint source’s unregulated pollution. For each $\zeta$, these aggregate pollutant levels can be interpreted as the imputed aggregate cap for each pollutant: these are the levels that would arise when trading is not allowed. The imputed caps are presented in Table 3 for each value of $\zeta$.

Comparing these values to the aggregate pollutant outcomes under various trading scenarios allows us to examine whether trading causes the imputed cap to be violated. The political viability of inter-pollutant trading could be in question if inter-pollutant trading increases pollutants in one medium relative to the no-trading outcome.

We now turn to the trading scenarios, starting with excessively strict caps ($\zeta = 0.95$). Consider the case of an integrated market with optimally-chosen intra-pollutant trading ratios (scenario 3). This scenario results in social net benefits that are 5 percent smaller than those in the efficient scenario due to over-regulation in scenario 3. Wastewater treatment plants bear the greatest costs of over-regulation. To offset these additional costs, the optimal inter-pollutant trade ratio ($\tau_{rW,el} = 0.014$) for scenario 3 is larger than the efficient ratio, reducing the incentives to reallocate abatement from industry to wastewater treatment facilities. Additionally, the intra-pollutant ratios ($\tau_{eA,el} = 0.318$ and $\tau_{rA,W} = 0.089$) are considerably smaller than the efficient values so that more abatement is allocated to nonpoint sources—including more exchange of industrial emissions for nonpoint loadings abatement. This further alleviates the burden of over-regulation on point sources and takes advantage of the complementarities in agricultural
abatement costs. The result is that agricultural abatement of both emissions and loadings increases by 26 percent relative to the efficient scenario, while point source abatement decreases relative to the efficient outcome. Relative to scenario 2, it is clear that the greatest reallocation of abatement responsibilities are from industrial emissions to agricultural loadings. Indeed, Table 3 indicates the implied loadings cap is satisfied for scenario 3, which stems from the significant increase in agricultural loadings abatement. However, the implied emissions cap is violated; since industry is by far the largest contributor of emissions, the 2 percent decrease in industry abatement in this case yields a significant increase in emissions. The intuitive reason for the imputed cap being violated is that the optimal market design is trying to move the system towards the efficient outcome, which involves greater pollutant levels in each sector.

Optimally-designed distinct markets under strict caps (scenario 4) are 5 percent less efficient than the corresponding integrated market (scenario 3) because distinct markets offer fewer instruments to reallocate abatement responsibilities. In particular, without the ability to directly exchange industrial emissions abatement for agricultural loadings abatement (the dominant form of exchange in scenario 3), the next best option is to reallocate emissions abatement from industrial sources to agricultural sources. This explains the intra-pollutant trading ratio for emissions being significantly smaller than in the integrated market scenario. The reallocation of emissions abatement to agricultural sources also results in significant loadings abatement by agricultural sources due to abatement cost complementarities, benefitting wastewater treatment plants. The intra-pollutant trading ratio for loadings increases in the distinct markets case to offset this effect. As with the case of integrated markets, the implied loadings cap is satisfied, whereas the implied emissions cap is violated (Table 3).
Further insight is obtained by examining the case where one-to-one intra-pollutant trade ratios are imposed. Consider first the integrated market scenario under strict caps (scenario 5). Social net benefits decline by 19 percent relative to scenario 3 because scenario 5 restricts the available tools for efficiently reallocating abatement responsibilities. The inter-pollutant trade ratio \( \tau_{W,eI} = \tau_{A,eI} = 0.0005 \) is smaller than any of the other ratios examined thus far, as this is now the only mechanism for reallocating abatement to the loadings sector. But note that, unlike the case with differentiated ratios, it is no longer possible to target reallocations towards abatement in agricultural loadings. Instead, the small inter-pollutant trade ratio encourages more loadings abatement by both agriculture and wastewater treatment plants. The result is greater control costs than in scenario 3. The imputed loadings cap is satisfied for this case, but the imputed emissions cap is violated.

The distinct market scenario 6 is even more limited in reallocating abatement efforts when the intra-pollutant ratios are fixed at unity because now policy makers have no tools available to guide the market. Consequently, this is the least efficient outcome overall. Social net benefits decline by 22 percent relative to scenario 5, which is a greater efficiency loss relative to the case when intra-pollutant trade ratios are chosen optimally; the inefficiency of eliminating inter-pollutant trading is compounded when the intra-pollutant trade ratios are restricted.

However, neither imputed cap is violated in this case due to the difficulty in encouraging reallocations across sectors.

The welfare rankings for the various scenarios involving excessively lax caps \( \zeta = 1.05 \) are the same as those for overly strict caps, whereas the optimal choices of trade ratios—and hence the allocation of abatement efforts—are largely opposite (Table 2), as might be expected. In contrast to trading under strict caps (scenarios 3–6), the imputed caps are not violated under
any scenario involving lax caps. The intuitive reason is that the optimal market design is trying
to move the system towards the efficient outcome, which involves smaller pollutant levels in
each sector. This result illustrates that trading—particularly inter-pollutant trading—is capable of
both reducing aggregate abatement costs and emissions relative to cases where trading is not
allowed. The importance of this result is magnified once we recognize that lax caps are more
likely to be applied in practice than strict caps.

Finally, Figure 2 illustrates social net benefits under each of the trading scenarios
described above over multiple values of $\zeta$. The figure shows the ranking of trade scenarios is
mostly preserved for other values of $\zeta$, with the welfare deviation expanding the further is $\zeta$ from
unity, which is the point where all scenarios converge at the efficient outcome. The only
exception is the integrated market with one-to-one inter-pollutant trading, which outperforms the
distinct markets with optimal inter-pollutant trading for large values of $\zeta$. Trading does not occur
in the emissions market for $\zeta > 1.04$; complementary emissions abatement by agriculture—
combined with industry’s own abatement—satisfies the emissions cap without need for trading.

6. Discussion and Conclusion

Growing recognition of the linkages between multiple pollutants has led economists and
regulatory agencies to consider multi-pollutant markets as a means to reduce the social cost of
environmental policies and improve efficiency (Lutter and Burtraw 2002; US EPA 2011). By
allowing trade across pollutants, multi-pollutant markets can exploit complementarities in
pollution abatement costs and the linkages between pollutant marginal damages, e.g., due to
cross-media damages.

Our findings reveal that in a world with no previous regulation in place and no
institutional constraints relative to setting the aggregate pollution cap, multi-pollutant markets for
uniformly mixed pollutants perform identically to distinct markets and replicate the first-best outcome in the absence of uncertainty. Hence, integrated markets for linked pollutants do not improve efficiency when caps can be set efficiently.

In practice, environmental managers engaged in permit market design decisions may face many practical constraints, including prior regulations that prevent permit caps from being set efficiently as a joint decision with the trade ratios. We demonstrate that such constraints have important consequences for market design choices in the context of multi-pollutant problems. In particular, it becomes optimal to choose trade ratios that cause the equi-marginal principle to break down in the permit market. This result contrasts with one basic tenet of environmental economics—that pollution markets minimize social costs by reallocating abatement effort to equate effective marginal abatement costs across sectors. Yet, the result stems from another basic tenet—the Tinbergen rule that one instrument is required for each externality. Specifically, the inability to set permit levels efficiently means that adjustments to the trade ratios, and an abandonment of the equi-marginal principle, are warranted to improve efficiency. This result holds even in the case of distinct markets when no inter-pollutant trading is allowed. However, as integrated markets involve an additional policy instrument that can ameliorate the efficiency losses from the permit restrictions (i.e., the inter-pollutant trade ratio), we find that integrated markets can be more efficient than distinct markets.

An important concern over integrated markets is that reallocating abatement effort across environmental media could, in principle, violate current environmental standards governing a particular medium. We find this is likely to be a concern when existing regulations are overly strict. However, in the more likely case that existing requirements are too lax relative to the efficient outcome (e.g., Muller and Mendelsohn 2009), we find that cross-media reallocations
can reduce social costs and improve environmental quality within each environmental medium. This is because trading ratios are optimally chosen to generate abatement closer to efficient levels. In cases where inter-pollutant trading does not reduce pollution universally, our results indicate that pollution is likely to decline for some media. More importantly, the overall economic efficiency of the solution will improve. Prior studies on air pollution (e.g., Lutter and Burtraw 2002) have argued for a need to focus on the efficiency gains arising from multi-pollutant trading, rather than focusing on whether suboptimal environmental standards are always satisfied.

A practical concern regarding multi-pollutant markets is the amount of information required to implement them. However, the only additional piece of information required relative to distinct markets is some measure of the complementarity in marginal abatement costs for linked pollutants, or the linkages in marginal damages across multiple media for pollutants affecting multiple media. Prior work empirically measures the joint production of multiple pollutants in the electrical utility industry (e.g., Agee et al. 2014), and doing so for other sectors is likely to be straightforward. It is more challenging to obtain estimates of the complementarity in marginal damages since, in many cases, the biogeochemical relationships between pollutants may not be well-known. However, as our understanding of the linkages between pollutants increases, so too will our ability to effectively manage them.

Finally, an important caveat of our results is that uncertainties stemming from stochastic emissions and uncertain abatement effectiveness, which are often characteristic of nonpoint pollution problems, were ignored here so that we could focus on the core issue of inter-pollutant trading. Incorporating these uncertainties into the analysis would be straightforward, however, and would mainly involve further modifications to the trade ratios. Recent numerical work on
nutrient water quality problems in the SRB (Horan and Shortle 2015) suggests that the
modifications could be positive or negative, depending on which type of uncertainty
(stochasticity, abatement effectiveness) dominates in the market equilibrium. We have also
ignored cost uncertainty, but including this feature would be unlikely to affect our qualitative
results since we compare market designs within a pollution permit trading framework rather than
compare different policy instruments (e.g., price and quantity instruments).

Appendix

Deriving Sector-Level Cost Minimization Problems

Consider the case of an integrated market. Following the approach of Horan and Shortle (2005),
suppose the wastewater sector may hold permits sold by any other sector. This sector’s initial
emissions permit holdings are \( \hat{r}_W \), and denote its purchases of permits from other sources by
\( \hat{e}_W, \hat{e}_A, \) and \( \hat{r}_A \). The sector’s costs of abatement and permit purchases are
\( C_W(a_W) + p_{eW} \hat{e}_W + p_{eA} \hat{e}_A + p_{eI} \hat{e}_I \). Moreover, this sector is constrained in that its total emissions cannot be greater
than its permit holdings, \( r_W \leq \hat{r}_W + \hat{e}_W/\tau_{el,W} + \hat{e}_A/\tau_{el,A} + \hat{r}_A/\tau_{el,A} \), where the final three
RHS terms represent the emissions that the wastewater treatment sector can generate based on
permits obtained from other sectors. Assuming the emissions constraint is satisfied as an
equality, then \( \hat{r}_A \) can be eliminated as a choice variable so that total costs are

\[ C_W(a_W) + p_{eA} \tau_{el,A} W \]

\[ = C_W(a_W) + p_{eA} \tau_{el,A} W \left[ r_W - \hat{r}_W - \hat{e}_W/\tau_{el,W} - \hat{e}_A/\tau_{el,A} \right] + p_{eI} \hat{e}_I + p_{eA} \hat{e}_A \]

Here, the choice variables are \( a_W, \hat{e}_W, \) and \( \hat{e}_A \). The FOCs associated with \( \hat{e}_W \) and \( \hat{e}_A \) are

\[ p_{el} = p_{eA} \frac{\tau_{el,A} W}{\tau_{el,W}} \]
where the equalities in equations (A2)–(A3) emerge in a competitive market equilibrium. Divide (A2) by (A3) to yield

\[
\frac{p_{e}}{p_{c}} = \frac{r_{e} - r_{c}}{r_{c} - r_{W}} = \frac{d\hat{a}_e}{d\hat{a}_c} = \tau_{e,c}.
\]

The cost minimization problems for the other sectors are analogous, with FOCs ultimately yielding the other two primary trade ratios

\[
\tau_{r,c} = \frac{p_{e}}{p_{r}}
\]

\[
\tau_{r,W} = \frac{p_{e}}{p_{r}}.
\]

Given these relations, the \(\hat{e}\) terms in (A1) cancel, so we can write the cost function for a particular sector restricted on only that sector’s pollution:

\[
C_j(a_{el}) + p_{el}[r_{l0} - a_{el} - \hat{e}_{l0}],
\]

\[
C_W(a_{rW}) + p_{rW}[r_{W0} - a_{rW} - \hat{r}_{W0}],
\]

and

\[
C_A(a_{el},a_{rA}) + p_{el}[e_{j0} - a_{el} - \hat{e}_{j0}] + p_{rA}[r_{j0} - a_{rA} - \hat{r}_{j0}].
\]

Deriving Second-Best Trade Ratios

Consider the integrated market, with the first-order conditions given by (17). In particular, the condition for the choice of \(\tau_{r,W,el}\) is

\[
\left[B'_{e} - C'_j \frac{\partial a_{el}}{\partial \tau_{r,W,el}} + B'_{e} - C'_{A} \frac{\partial a_{el}}{\partial \tau_{r,W,el}} + B'_{r} - C'_{r} \frac{\partial a_{rW}}{\partial \tau_{r,W,el}} \right] = \frac{\partial a_{rW}}{\partial \tau_{r,W,el}} = 0
\]

which can be rearranged and written as
the term depends on all trade ratios. Likewise, we can derive

\[ \tau_{eA,el} = \frac{B'_e}{\partial C'_A / \partial a_{eA}} \rho_{el}(\tau_{eA,el}) + \left[ \frac{B'_e - \partial C'_A / \partial a_{eA}}{\partial C'_A / \partial a_{eA}} \right] \rho_{el}(\tau_{eA,el}) \]

\[ + \frac{B'_r}{\partial C'_A / \partial a_{eA}} \left[ \rho_{el}(\tau_{eA,el}) + \rho_{el}(\tau_{eA,el}) \right] \]

\[ \tau_{rA,el} = \frac{B'_e}{\partial C'_A / \partial a_{eA}} \left[ \rho_{el}(\tau_{rA,el}) + \rho_{el}(\tau_{rA,el}) \right] + \left[ \frac{B'_e - \partial C'_A / \partial a_{eA}}{\partial C'_A / \partial a_{eA}} \right] \rho_{el}(\tau_{rA,el}) \]

\[ + \frac{B'_r}{\partial C'_A / \partial a_{eA}} \rho_{el}(\tau_{rA,el}) \]
Now consider the case of distinct markets. The optimality conditions (17) still hold. Using a process analogous to that described above, the optimal trade ratios are (A10) and 

\[
\tau_{r, l,W} = \frac{B'_r}{\partial C_A/\partial a_{r, l}} \rho_{l,W}(\tau_{r, l,W}) + \frac{B'_r - \partial C_A/\partial a_{r, l}}{\partial C_A/\partial a_{r, l}} \rho_{r, l}(\tau_{r, l,W}) \\
+ \frac{B'_r}{\partial C_A/\partial a_{r, l}} \left[ \rho_{cl}(\tau_{r, l,W}) + \rho_{el}(\tau_{r, l,W}) \right].
\]

(A12)

References


Table 1. Simulation Parameters

<table>
<thead>
<tr>
<th>Sector</th>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial</td>
<td>$\psi$</td>
<td>Marginal cost parameter, industrial sector</td>
<td>$8.93 \times 10^4$</td>
</tr>
<tr>
<td></td>
<td>$e_{I0}$</td>
<td>Initial industrial emissions (million mtCO2e)</td>
<td>153.5</td>
</tr>
<tr>
<td>Wastewater</td>
<td>$\phi$</td>
<td>Marginal cost parameter, wastewater treatment sector</td>
<td>$6.4 \times 10^6$</td>
</tr>
<tr>
<td></td>
<td>$r_{W0}$</td>
<td>Initial wastewater treatment loadings (thousand mtN)</td>
<td>12</td>
</tr>
<tr>
<td>Agricultural</td>
<td>$\alpha$</td>
<td>Marginal cost parameter, agricultural sector</td>
<td>$2.08 \times 10^8$</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>Marginal cost parameter, agricultural sector</td>
<td>$3.71 \times 10^5$</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>Complementarity parameter, agricultural sector</td>
<td>$1.1 \times 10^7$</td>
</tr>
<tr>
<td></td>
<td>$e_{A0}$</td>
<td>Initial agricultural emissions (million mtCO2e)</td>
<td>5.54</td>
</tr>
<tr>
<td></td>
<td>$r_{A0}$</td>
<td>Initial agricultural loadings (thousand mtN)</td>
<td>28</td>
</tr>
<tr>
<td>Damages</td>
<td>$\varepsilon$</td>
<td>Marginal damage from emissions</td>
<td>$1.4 \times 10^7$</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>Marginal damage from loadings</td>
<td>$1.65 \times 10^5$</td>
</tr>
</tbody>
</table>
Table 2. Simulation Results from Pollutant Trading Scenarios

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Social net benefits ($ million)</th>
<th>Trading ratios</th>
<th>Emissions abatement (million mtCO(_2)e)</th>
<th>Loadings abatement (1,000 mtN)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(\tau_{\text{W,eI}})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Efficient (first-best)</td>
<td>163.4</td>
<td>0.0025</td>
<td>1</td>
<td>0.61</td>
</tr>
<tr>
<td>2. Exogenous caps: strict ((\zeta = 0.95))</td>
<td></td>
<td>0.0005</td>
<td>1</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.012</td>
<td>12.11</td>
<td>0.77</td>
</tr>
<tr>
<td>3. Optimally chosen trade ratios</td>
<td></td>
<td></td>
<td>12.29</td>
<td>0.39</td>
</tr>
<tr>
<td>4. One-to-one intra-pollutant ratios</td>
<td></td>
<td></td>
<td>11.83</td>
<td>0.74</td>
</tr>
<tr>
<td>5. Exogenous caps: lax ((\zeta = 1.05))</td>
<td></td>
<td></td>
<td>6.11</td>
<td>6.39</td>
</tr>
<tr>
<td>6. Optimally chosen trade ratios</td>
<td></td>
<td></td>
<td>11.22</td>
<td>1.58</td>
</tr>
<tr>
<td>7. One-to-one intra-pollutant ratios</td>
<td></td>
<td></td>
<td>11.22</td>
<td>5.02</td>
</tr>
<tr>
<td>8. Exogenous caps: lax ((\zeta = 1.05))</td>
<td></td>
<td></td>
<td>6.11</td>
<td>6.39</td>
</tr>
<tr>
<td>9. Optimally chosen trade ratios</td>
<td></td>
<td></td>
<td>5.02</td>
<td></td>
</tr>
<tr>
<td>10. One-to-one intra-pollutant ratios</td>
<td></td>
<td></td>
<td>5.02</td>
<td></td>
</tr>
</tbody>
</table>

\(\text{WWT}^a\) = Wastewater treatment sector.  
\(\text{Emissions trading breaks down for a lax cap.}\)
<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Emissions (million mtCO\textsubscript{2}e)</th>
<th>Loadings (thousand mtN)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Imputed cap</td>
<td>Total emissions</td>
</tr>
<tr>
<td>Exogenous caps: strict ($\zeta = 0.95$)</td>
<td>138.89</td>
<td></td>
</tr>
<tr>
<td><em>Optimally chosen trade ratios</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Integrated market</td>
<td>145.98</td>
<td></td>
</tr>
<tr>
<td>4. Distinct markets</td>
<td>146.0</td>
<td></td>
</tr>
<tr>
<td><em>One-to-one intra-pollutant ratios</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Integrated market</td>
<td>146.47</td>
<td></td>
</tr>
<tr>
<td>6. Distinct markets</td>
<td>138.89</td>
<td></td>
</tr>
<tr>
<td>Exogenous caps: lax ($\zeta = 1.05$)</td>
<td>152.93</td>
<td></td>
</tr>
<tr>
<td><em>Optimally chosen trade ratios</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Integrated market</td>
<td>147.28</td>
<td></td>
</tr>
<tr>
<td>9. Distinct markets</td>
<td>152.32</td>
<td></td>
</tr>
<tr>
<td><em>One-to-one intra-pollutant ratios</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Integrated market</td>
<td>152.22</td>
<td></td>
</tr>
<tr>
<td>11. Distinct markets</td>
<td>152.93</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Shaded cells highlight trading scenarios that do not violate the imputed cap.
Figure 1. The Susquehanna River Basin
Figure 2. Social net benefits from abatement in the SRB under separate and integrated markets when permit caps set exogenously at the fraction $\zeta_r = \zeta_e = \zeta$ of efficient levels.