The Economics of Enforcing Emissions Markets: A Review of the Literature

JOHN K. STRANLUND

Department of Resource Economics
University of Massachusetts-Amherst

September 5, 2015

Correspondence: John K. Stranlund, Department of Resource Economics, 222 Stockbridge Hall, 80 Campus Center Way, University of Massachusetts-Amherst, Amherst, MA 01003, USA. Telephone: (413)-545 6328, E-mail: stranlund@resecon.umass.edu.

JEL Codes: K42, L51, Q58

Introduction

Despite the advantages of market-based environmental policies over traditional command-and-control approaches, these programs are not likely to perform as expected if they are not enforced well. Accordingly, there is now a significant body of literature on the economics of compliance and enforcement in emissions trading programs. The purpose of this article is to review this literature. There are several other reviews of the economic literature on enforcing environmental policies, including Cohen (1999), Heyes (2000), Gray and Shimshack (2011), and Shimshack (2014); however, no one has provided a thorough review of the literature that focuses on enforcing emissions markets.
Much of the economic literature on enforcing emissions markets consists of theoretical models developed to derive positive results about firms’ compliance behavior and the effects of enforcement on the performance of emissions markets. Empirical tests of these results have been limited because of the lack of useful field data. In situations in which field data are unavailable laboratory experiments can provide valuable information, and several authors have used these experiments to test hypotheses related to compliance in emissions markets. Positive theoretical and experimental results have also been used to draw normative conclusions about designing efficient enforcement strategies. In principle the enforcement component of a trading program should be determined jointly with all its other elements. The literature on enforcing emissions markets provides insights into the efficient design of these programs, including their enforcement strategies.

The article proceeds as follows. The next section is a very brief overview of some of the enforcement strategies used in existing and proposed emissions markets to provide some institutional context. I then review fundamental theoretical hypotheses about source compliance and enforcement, as well as the results of laboratory tests of these hypotheses. This section highlights the important role that permit prices play in determining compliance choices and the productivity of enforcement strategies. This section also examines how enforcement affects the cost-effectiveness of emissions markets, as well as the compliance issues arising from market power and transaction costs. Next I review the literature on the efficient design of enforcement strategies for emissions markets. I first consider the problem of enforcing dynamic emissions markets, which brings out the critical importance of self-reported compliance data. This section then turns to the question of whether emissions markets should be designed to motivate full compliance, or whether permitting a certain amount of noncompliance can reduce the costs of
reaching environmental quality goals. This discussion leads to one about the normative relationship between enforcement and provisions to deal with uncertainty about firms’ abatement costs. In the penultimate section I use results from the literature to draw lessons about enforcement and compliance in international markets for greenhouse gas control, including the enforcement consequences of linking regional markets together and the use of offsets. I conclude in the final section.

**Institutional Context**

The fundamental problem of enforcing emissions markets is to make sure that pollution sources hold enough permits to cover their current emissions. This requires that regulators have systems in place to track emissions permits so that they know how many permits a source holds at any point in time. The much more difficult problem, however, is to monitor sources’ emissions. This can be done directly with the use of continuous emissions monitoring systems, which measure the continuous flow of pollution leaving a facility. The largest sources in the US EPA’s SO₂ Allowance Trading Program and NOₓ Budget Trading Program were required to install these systems (US EPA 2009a). An alternative is to allow sources to estimate their emissions. Smaller sources in the EPA’s air markets as well as nearly all sources in the European Union’s Emissions Trading Scheme (EU ETS) estimate their emissions rather than monitor them directly. Estimates tend to be based on formulae that combine activity data like fuel and raw material use with emissions factors that specify emissions per unit of the activity (Kruger, Oates and Pizer 2007, McAllister 2010).

As with most environmental law, emissions markets rely heavily on self-monitoring and self-reporting of data used to determine compliance. This data includes emissions or estimated
emissions and the data used for the estimates, as well as quality assurance and quality control information related to the operation of monitoring technologies. Some programs, like the EU ETS and California’s greenhouse gas market require that firms’ emissions reports be verified by a certified, independent auditor (Fleurke and Verschuuren 2015, California Air Resources Board 2011)

Comparing a firm’s permit holdings to its emissions (or estimated emissions) for a compliance period determines whether it is violating its permits. Sources that have excess emissions face sanctions that vary a great deal across emissions markets. Some programs sanction permit violations with an automatic financial penalty and a surrender of permits to offset the shortfall. For example, the permit violation penalty in the SO2 Allowance Trading Program was set at $2,000 per ton of excess emissions in 1990, and was adjusted for inflation every year. In addition, excess emissions are offset from allocations in the next year (US EPA 2010). Similarly, the EU ETS includes a €100 per ton permit violation penalty and excess emissions must be offset in the following year (Fleurke and Verschuuren 2015). Other programs use permit offsets as the main penalty. For example, permit violations in the NOx Budget Trading Program were penalized with a three-to-one offset (US EPA 2009b). The model rule for the Regional Greenhouse Gas Initiative also includes this penalty (Regional Greenhouse Gas Initiative 2013). Similarly, permit violations under California’s greenhouse gas trading program are punished with a four-to-one offset (California Air Resources Board 2011). In nearly all trading programs the value of permit violation sanctions is substantially higher than prevailing permit prices.

A little-used but potentially important form of sanction ties a financial penalty directly to prevailing permit prices. The proposed (but not enacted) Clear Skies Initiative in the US included
a permit violation penalty that was to be one to three times the clearing price in a recent permit auction (US EPA 2003). Similarly, the American Clean Energy and Security Act of 2009 (US Congress 2009) included a permit violation penalty that was set at twice the market value of CO2 permits during a compliance year. Perhaps the only example of this form of sanction in actual practice occurs in pilot programs for China’s CO2 emissions trading system. In the Beijing and Shenzhen pilot programs, penalties for excess emissions are set at three to five times the market price of CO2 allowances (Peeters and Chen 2015). In this review we will see how this form of penalty can shield compliance choices and enforcement effort from variation in permit prices.

While much attention has been given to the role of high permit violation penalties for maintaining compliance in emissions markets, penalties for reporting violations can also be significant. For example, the US Clean Air Act authorizes civil and criminal sanctions for false reporting under the EPA’s emissions markets (Tietenberg 2006, McAllister 2010). Sanctions for reporting violations in the EU ETS are left to member states. There is a lot of variation in these penalties, but they include both financial sanctions and the threat of prison terms (European Environment Agency 2008). We will see later that the main challenge of maintaining compliance in emissions markets may actually lie in motivating accurate reporting.

The performance of enforcement strategies for emissions markets has ranged from programs with significant noncompliance as well as examples with near-perfect compliance. Montero et al. (2002) and Palacios and Chavez (2005) argue that the development of an emissions trading program for suspended particulates in Santiago, Chile was hampered by weak enforcement and significant noncompliance. McAllister (2010) reports on compliance problems in the RECLAIM program. On the other hand, the US EPA’s air markets have achieved very high rates of compliance (US EPA 2009b, 2010).
Enforcement and Compliance in Emissions Markets

This section reviews fundamental, largely positive, results about enforcement in emissions markets. I start with results about compliance choices and enforcement-induced market effects from static models of competitive emissions markets. I then consider how the enforcement problem affects the cost-effectiveness of emissions markets, and the impacts of market imperfections. While much of this section is focused on positive results, these often lead directly to normative lessons about how best to design enforcement strategies.

I also review empirical tests of many of the theoretical results presented in this section. These tests are limited to studies with laboratory experiments, because opportunities for using field data are limited. Programs that have achieved very high rates of compliance have too little variation in compliance choices to conduct meaningful statistical tests, and there is little analysis of compliance behavior in programs that have had significant noncompliance.

Fundamentals of compliance in emissions markets

In a simple static model of compliance under a competitive emissions market, a risk-neutral firm chooses its emissions and the number of permits to hold to minimize its costs of abatement plus its expenditures or receipts on permit transactions, and its expected penalty should it decide not to hold enough permits to cover its emissions. The expected penalty consists of a probability that the firm is inspected and a violation is discovered and a financial sanction that is increasing and convex in the size of the firm’s violation. All of the theoretical results in this review assume that firms choose a positive amount of emissions and hold a positive number of permits.

1 The literature that uses laboratory experiments to evaluate various aspects of emissions markets is extensive. See Friesen and Gangadharan (2013) for a recent review of this literature.
Under these conditions a firm holds enough permits to cover its emissions if and only if the prevailing permit price is not less than the expected marginal penalty of even a small violation. Moreover, if the firm decides to violate its permits it chooses a violation level so that the permit price is equal to the expected marginal penalty (Malik 1990, Keeler 1991). The permit price plays a key role in compliance decisions and enforcement because it is the common observable marginal benefit of firms’ noncompliance. In addition, given reasonable assumptions about the monitoring probability, a firm’s choice of emissions follows the familiar rule of equating the permit price to its marginal abatement costs. ²

Stranlund and Dhanda (1999) point out that since firms’ compliance decisions depend only on the permit price and the enforcement strategy they face, differences in their compliance choices are independent of differences in their abatement costs and initial permit allocations. Conceptually, there is no reason for regulators to believe that some firms will be more likely to be noncompliant or tend toward higher violations even though they may have different abatement or production technologies, or initial permit allocations. This implies that enforcers cannot make their efforts more productive by targeting firms (perhaps with differential monitoring) based on differences in their observable characteristics. This simplifies enforcement, because it means that regulators do not need information about firms’ characteristics that may be difficult to obtain to help them devise an effective enforcement strategy—they only need to observe the permit price.

Matters are very different for firms that face command-and-control standards. For example, a firm’s decision about whether to comply with an emissions standard is determined by the relationship between its marginal abatement cost and the marginal expected penalty it faces.

² In addition, all of the theoretical results in this review assume that firms choose a positive amount of emissions and hold a positive number of permits.
Consequently, firm-level characteristics are important determinants of compliance with fixed standards (Garvie and Keeler 1994). Empirical support for this conclusion is provided by Gray and Shadbegian (2005 and 2007) and Earnhart (2009). Since firms’ abatement costs partly determine their compliance choices when they face emissions standards, authorities can condition their enforcement effort on observable firm characteristics that are correlated with their abatement costs to help achieve compliance goals. This would not be productive under a competitive emissions market.

Murphy and Stranlund (2007) provide experimental support for these conclusions. They found in their laboratory markets that subjects’ violations were independent of differences in their abatement costs, as predicted. However, they also found that subjects who were predicted to buy permits tended to have higher violations than those who were predicted to sell permits. While this suggests that enforcers may be motivated to target permit buyers with more intense monitoring, Murphy and Stranlund (2007) found that the marginal productivity of increased enforcement in reducing violations was independent of differences in both abatement costs and initial permit allocations. Thus, under the important objective of maximizing aggregate compliance with scarce enforcement resources, regulators have no theoretical or empirical justification for targeting firms based on their individual characteristics.

Another consequence of the key role that permit prices play in determining compliance incentives is that changes in the regulatory environment can have direct effects on compliance choices, as well as indirect effects through changes in the permit price. For example, increased monitoring or higher sanctions leads noncompliant firms to purchase more permits to reduce

---

3 I use standard environmental policy language to describe the results of laboratory emissions markets, but these experiments typically use neutral terminology that avoids placing subjects in an explicit environmental policy setting.
their violations. This is the direct effect of enforcement on compliance. However, the increased demand for permits can increase the permit price, which motivates firms to higher violations. This is the indirect effect of increased enforcement on compliance that works in the opposite direction as the direct effect. Theoretically, the direct effect always outweighs the indirect effect so greater enforcement produces greater compliance, but regulators need to be aware that the productivity of enforcement pressure in reducing noncompliance is partially offset by a countervailing price effect. Moreover, that a firm chooses its emissions to equate its marginal abatement costs to the permit price, whether it is compliant or not, implies that this choice does not depend directly on the enforcement strategy it faces.—emissions fall only if increased enforcement leads to a higher permit price. Murphy and Stranlund (2006) confirmed these hypotheses about the direct and indirect price effects of enforcement in laboratory markets.

The special features of compliance decisions in emissions markets have important implications for the adoption and diffusion of new cost-saving abatement technologies. Imagine a firm that violates a fixed emissions standard. If it adopts a new abatement technology it will reduce its violation because its marginal abatement costs fall. This is not true of a firm in an emissions market. Arguedas, Camacho and Zofio (2010) and Villegas-Palacios and Coria (2010) show that since firms’ violation choices do not depend on their abatement costs, a firm that adopts a new abatement technology will not change its violation level if the permit price is unchanged. However, if the technology is diffused enough to lower the permit price, then all firms, not only those that adopt the new technology, will reduce their violations. This suggests that policies to promote cost-saving abatement technologies can also reduce firms’ incentive to violate their permits, but only if the policies lead to lower permit prices.
Enforcement and cost-effectiveness

Malik (1990) appears to be the first to show that under reasonable assumptions about the structure of a fixed enforcement strategy, each firm in a competitive emissions market chooses its emissions to equate its marginal abatement costs to the permit price, whether it is compliant or not. Of course, this implies that aggregate abatement costs are minimized even though aggregate emissions may exceed the permit cap because of weak enforcement. Stranlund, Murphy and Spraggon (2013) tested this hypothesis in laboratory markets and found that these markets were very close to minimizing aggregate abatement costs, despite imperfect enforcement and significant violations.4

While we might expect a competitive permit market to minimize aggregate abatement costs, even with weak enforcement, Malik (1992) demonstrated that an emissions market will not, in general, minimize the sum of abatement costs and enforcement costs. The problem, according to Stranlund, Chavez and Villena (2009), is that competitive permit trading typically leads to a uniform price and minimum aggregate abatement costs, while the addition of enforcement costs may call for discriminatory prices that reflect differences in marginal enforcement costs across firms. These differences may be due to differences in the ability of regulators to monitor firms’ emissions, as well as the slopes of firms’ abatement costs that determine their responsiveness to additional enforcement pressure. Implementation of discriminatory prices in emissions markets can be accomplished with permit trading ratios among firms that differ from conventional one-for-one trading schemes.

4 They also found that aggregate violations and emissions tended to be significantly lower than predicted. This is a common result that has been observed in many settings including other emissions trading experiments (e.g., Raymond and Cason 2011).
Whether discriminatory pricing to account for differences in marginal enforcement costs is worthwhile depends on whether regulators have enough information to detect these differences and whether they are significant enough to justify the added complexity. At present, I know of no proposal to implement discriminatory emissions prices to account for differences in enforcement costs. The rest of this review focuses on enforcing markets that feature uniform permit prices.

**Market imperfections**

The theory and experiments discussed thus far were constructed under the assumption of perfect competition. There are several sources of market imperfections that can lead permit markets to inefficient outcomes, including the presence of market power and transaction costs. These imperfections can have important implications for compliance behavior and enforcement as well.

van Egteren and Weber (1996), Malik (2002) and Chavez and Stranlund (2003) each extended Hahn’s (1984) seminal paper on market power in emissions markets to account for compliance and enforcement. This literature shows that the presence of market power can affect the compliance incentives of all firms. As a powerful permit buyer limits the number of permits it buys to hold the permit price down, it reduces the incentive for firms in a competitive fringe to violate their permits. At the same time, the dominant firm’s marginal benefit of noncompliance is higher than that of the competitive firms, because it is the sum of the going price and the marginal increase in this price as the firm buys additional permits. These effects are reversed if the dominant firm sells permits. Recall that enforcement of a perfectly competitive market is made simpler because regulators have no reason to target enforcement at different firms and they do not need information about firms’ abatement costs to help them devise effective enforcement. These nice features disappear in the presence of powerful firms. Since their compliance
incentives are different from the incentives of the competitive fringe, regulators have some incentive to target powerful firms with more or less enforcement pressure, depending on whether the firms are permit buyer or sellers.

Transactions costs can also have a significant impact on compliance behavior and enforcement of trading programs. Stavins (1995) showed that variable transaction costs drive a wedge between the price that buyers pay for permits and the price that sellers receive (also see Montero 1998). Consequently, transaction costs can also drive a wedge between the compliance incentives of buyers and sellers; in particular, permit buyers are more likely to violate their permits than sellers. Like market power, transaction costs associated with permit trading can make the problem of enforcing emissions markets more complex.

Cost-Effective and Efficient Enforcement

Thus far, this review has focused largely (but not entirely) on positive compliance and market results under the assumption that the parameters of an emissions trading program (monitoring, penalties, and permit supply) are exogenous. We now turn exclusively to the design of enforcement strategies. I begin with the problem of enforcing dynamic emissions markets, which highlights the critical role played by self-reporting of compliance data. I then turn to the joint determination of enforcement along with the other components of a trading program, focusing on the optimal level of noncompliance in these programs and the interaction between enforcement and strategies to address uncertainty in firms’ abatement costs. All of the literature reviewed in this section assumes that permit markets are perfectly competitive.
Intertemporal permit trading and the role of self-reporting

Most emissions markets allow firms the limited ability to bank emissions permits. For example, the SO$_2$ Allowance Trading Program allowed firms to save permits for future use or sale, but did not allow them to borrow against future permit allocations. The NO$_x$ Budget Trading Program had similar banking provisions, except it imposed a heavy discount on saved permits if the aggregate bank reached a particular limit (US EPA 2009b). Some programs for greenhouse gas emissions allow restricted permit borrowing as well. Examples include the EU ETS and the proposed American Clean Energy and Security Act of 2009 (US Congress 2009). Permit banking allows firms to shift abatement across time cost-effectively and to hedge against risk associated with uncertain abatement costs, emissions, and permit prices.

Innes (2003) was the first to examine the compliance problem in a model of intertemporal permit trading. In his model, permit banking is motivated by stochastic emissions and enforcement is as simple as possible; a regulator can observe each firm’s emissions perfectly and without cost, but it is costly to impose penalties for permit violations. Firms that have imperfect control of their emissions can find themselves short of permits at the end of a compliance period, but Innes demonstrated that allowing intertemporal permit trading can eliminate costly sanctions because it allows firms to borrow against future permit allocations to make up a current permit shortfall. Moreover, firms may save permits as a hedge against permit shortfalls in the future, and firms that hold excess permits for this purpose are obviously compliant in the current period.

While Innes (2003) assumed a very simple enforcement apparatus, Stranlund, Costello, and Chavez (2005) examined the design of monitoring and sanctioning strategies in dynamic trading environments. Their most important contribution is to highlight the importance of self-reported compliance data. They note first that the combination of imperfect emissions
monitoring and bankable permits requires that firms self-report their emissions. The reason is that if a firm is not monitored in a period, its emissions report is the only information available to determine how many permits are used for current compliance purposes and how many are carried into the future. Moreover, misreporting and the failure to hold sufficient permits must be distinct violations. This is so because a firm that holds enough permits to cover its emissions in a period may still be motivated to underreport its emissions to increase the size of its permit bank.\(^5\)

Stranlund et al. (2005) also demonstrate that the high permit violation penalties (relative to permit prices) that are characteristic of many emissions trading programs have little deterrence value. The strong incentive to bank permits and the common requirement to offset any permit violation with a reduction in a future permit allocation effectively eliminates the incentive to violate one’s permits. In principle, permit violation penalties need to only cover the difference between this period’s permit price and the present value of next period’s price, and hence, can normally be very low. In addition, setting a high permit violation penalty cannot reduce monitoring effort. In contrast, a specific penalty for underreported emissions allows regulators to maintain compliance with imperfect monitoring, and setting this penalty as high as is practicable conserves monitoring costs.

Stranlund, Murphy and Spraggon (2011) examined compliance and banking in laboratory emissions markets to test these hypotheses. Subjects in their experiments were motivated to bank permits to smooth a decrease in the aggregate supply of permits in the middle of multi-period trading sessions. One of their treatments was parameterized to induce full compliance according

---

\(^5\) Requiring self-reporting and making misreporting a distinct violation differs fundamentally from self-discovery and disclosure rules that seek to encourage greater compliance with environmental regulations by reducing penalties for violations that are voluntarily discovered and reported to authorities. For an example, see the US EPA’s Audit Policy (US EPA 2000). Interestingly, much of the economic literature on self-reporting in law enforcement assumes that self-reporting is a voluntary activity that can be encouraged by offering a lower penalty for self-reported violations (e.g., Malik 1993, Kaplow and Shavell 1994, and Innes 2001).
the theoretical model of Stranlund et al. (2005), but with reporting and permit violation penalties set below expected permit prices. In fact, the permit violation penalty was set at about one-quarter of the predicted price. Reporting and permit compliance rates in this treatment were quite high, about 96 percent and 92 percent compliance, respectively. Stranlund et al. (2011) conducted an additional treatment that reduced the monitoring probability by half to investigate the consequences of weak enforcement on dynamic emissions markets. As expected, there was significant noncompliance in this treatment, but nearly all of it involved reporting violations: permit compliance in this treatment remained high. These results support the theoretical notions that high permit violation penalties have little deterrence value in markets with bankable permits, and that the main enforcement challenge in these environments is to motivate accurate self-reporting.

Cason and Gangadharan (2006) also studied compliance and banking in a laboratory market, but they focused on how the ability to bank permits affects compliance, not on the design of enforcement strategies given the ability to bank permits. Banking in their experiments was motivated by giving subjects imperfect control over their emissions. Surprisingly, they found that banking led to significant noncompliance and higher emissions. Although they did not distinguish between reporting and permit violations in their data analysis, it appears that most of the violations were reporting violations. In practice, the opportunity firms have to increase the size of their permit banks by underreporting current emissions may increase their motivation to submit false emissions reports. However, it is clear that enforcement strategies can be designed to deter this behavior.

Another feature of dynamic emissions markets has enforcement consequences; that is, as prices evolve over time so too do firms’ compliance incentives. For example, banking permits
under certainty requires that the nominal price of permits rises at the rate of discount (Rubin 1996), which increases the incentive for sources to underreport their emissions. To counteract this effect, either sanctions must increase or monitoring intensity must increase with the permit price. Since increasing monitoring is costly but increasing penalties typically is not, sanctions should respond to price increases, not the level of monitoring. A simple way to do this is to make misreporting sanctions a constant multiple of realized permit prices. This keeps monitoring effort and compliance choices constant as permit prices increase. Recall that tying sanctions to permit prices has been proposed in two U.S. trading programs and has been implemented in pilot trading programs in China.

**Cost-effective enforcement**

A key issue in the design of any environmental policy is whether they should be designed to motivate full compliance, or whether permitting a certain amount of noncompliance can reduce the costs of reaching environmental quality goals. Much of literature on designing emissions markets sidesteps this issue by assuming that regulators cannot apply enough enforcement pressure to induce compliance (e.g., Malik 1990 and 2002, Keeler 1991, van Egteren and Weber 1996, Villegas-Palacios and Coria 2010). For example, Stranlund and Dhanda (1999) examined the strategy of an enforcer who seeks to minimize noncompliance with fixed budget that is not large enough to achieve full compliance. While limited enforcement resources are certainly a factor in many instances of regulatory enforcement, an enforcer’s budget is an endogenous element in the determination of an optimal emissions market. Another common assumption to preclude full compliance is that penalties are restricted. For example, Montero (2002) assumes that permit violation penalties cannot be set above permit prices. This assumption is overly
restrictive and uncharacteristic of real emissions markets in which penalties are often many times higher than prevailing permits prices. Other authors assume away the problem of the optimal level of noncompliance by restricting their analyses to full-compliance outcomes without justifying this choice from an efficiency standpoint (Malik 1992, Chavez and Stranlund 2003, Stranlund et al. 2005). In practice, recall that there are examples of emissions markets with significant noncompliance as well as examples with nearly perfect compliance.

Stranlund (2007) addressed the problem of the optimal level of noncompliance in an emissions market. In his static model a regulator chooses a supply of emissions permits and an enforcement strategy to minimize the expected costs of achieving a fixed aggregate emissions limit. (Caffera and Chavez (2011) provide additional results about the cost-effectiveness of alternative levels of noncompliance for both emissions standards and markets). Expected costs include the firms’ abatement costs and the government’s monitoring costs, and also the expected costs of sanctioning noncompliant firms. Modeling sanctioning costs is not common in the literature on enforcing environmental policies (although Malik (1993), Arguedas (2008) and Stranlund et al. 2009 are other exceptions). In reality, however, penalizing firms may include the administrative costs associated with collecting penalties, the costs of investigative efforts, the costs of firms’ efforts to challenge or avoid the imposition of penalties, and the government’s costs of confronting these challenges.

Stranlund (2007) demonstrated that any static emissions trading scheme that achieves an aggregate emissions goal while tolerating permit violations is more expensive than an alternative policy that achieves the same aggregate emissions, but motivates firms to be fully compliant. The keys to this results are, firstly, motivating full compliance eliminates variable sanctioning costs, and secondly, there are sufficient levers in the design of a trading program (the permit supply,
monitoring, and penalty function) to achieve any level of aggregate emissions with full compliance without intensifying monitoring effort or setting higher marginal penalties. For a simple example of this principle, suppose that the expected marginal penalty for permit violations is a constant that is applied to all firms and is not high enough to induce full compliance with a fixed number of permits. In this case the equilibrium price of permits is equal to the common expected marginal penalty. A simple alternative is to increase the supply of permits by the amount of aggregate violations without changing anything else about the policy. This modification leads to the same outcome in terms of equilibrium price and aggregate emissions, but noncompliance and variable sanctioning costs are eliminated, thereby reducing enforcement costs without incurring other costs.

The result that it is cost-effective to design emissions markets that achieve full compliance extends to designing an efficient policy if firms’ abatement costs are known with certainty. The reason is that motivating full compliance is a cost-minimizing strategy for achieving any given level of aggregate emissions, including the one that balances the costs and benefits of emission control efficiently. However, the issue of the optimal amount of noncompliance is more complicated when firm’s abatement costs are uncertain.

**Abatement cost uncertainty**

Policy design under uncertainty about firms’ abatement costs has pre-occupied environmental economists for many decades. Weitzman’s (1974) derivation of rules for choosing between price and quantity policies continues to be relevant in today’s debates. For example, it is widely believed that a carbon tax is more efficient than a carbon market, because the marginal damage of carbon emissions is almost perfectly flat over a compliance period (Pizer 2002). However, the
preference in many circles for markets over taxes has generated rapid innovation in hybrid schemes. The most popular form of these hybrids involve tradable permits with price controls of as first proposed by Roberts and Spence (1976). Many recent policies to control greenhouse gases include some form of price control. See Hood (2010) and Newell, Pizer and Raimi (2013) for several examples.

The first proposals for hybrid schemes for carbon markets involved price ceilings, implemented by allowing firms to purchase additional permits from the government at a pre-determined price (Pizer 2002, Jacoby and Ellerman 2004). These policies are also known as safety valves, because they allow firms to escape the limit imposed by the permit supply in case their abatement costs turn out to be significantly higher than expected. Adding a price floor can improve efficiency by motivating firms to abate below the permit supply if their abatement costs turn out to be lower than expected. Several recent simulation studies demonstrate the cost-effectiveness of combining price ceilings and price floors (Burtraw, Palmer and Kahn 2010, Fell et al. 2012).

Constructing hybrid control policies has an enforcement angle, because some authors view the high penalties for permit violations in the SO2 Allowance Trading program and the EU ETS as price ceilings that can limit high-side abatement cost risk (Jacoby and Ellerman 2004, Stavins 2008). More rigorously, Montero (2002) reexamined the prices vs. quantities debate to analyze the effects of imperfect and costly enforcement on the choice between an emissions tax and emissions trading. He found that imperfect enforcement tends to favor emissions trading precisely because the expected marginal penalty can provide the price ceiling that improves the efficiency of emissions trading under abatement cost uncertainty. If firms’ abatement costs turn out to be very high, the permit price will rise to the expected marginal penalty and firms would
increase their emissions beyond the permitted cap by violating their permits. Thus, uncertainty about firms’ abatement costs can provide a justification for designing emission markets that may result in imperfect compliance when firms’ abatement costs are higher than expected.

Stranlund and Moffitt (2014) note several problems with using imperfect enforcement in this way. First, sanctioning noncompliant firms is costly, so this scheme involves the expectation of incurring sanctioning costs when firms’ abatement costs are high. Second, imperfect enforcement allows the transmission of abatement cost risk to compliance choices and enforcement costs. Third, if expected sanctioning costs are an increasing function of expected sanctions, then the optimal trading program that uses noncompliance as a safety valve involves weaker emissions control (in expectation). Finally, imperfect enforcement cannot be used to address low-side abatement cost risk, because it cannot motivate extra abatement when abatement costs turn out to be lower than expected. Stranlund and Moffitt (2014) show how each of these costly features are eliminated with a perfectly enforced trading program that features an explicit price ceiling and floor, as well as sanctions that vary directly with the permit price. The policy addresses both low-side and high-side abatement cost risk; full compliance eliminates variable sanctioning costs, and making sanctions vary with the permit price prevents the transmission of abatement cost risk to enforcement costs.

Lessons for Greenhouse Gas Trading

While the literature on the economics of enforcing emissions markets has progressed quite far, little of it has focused specifically on the control of markets for greenhouse gas emissions. In this section I draw six lessons from the literature that apply generally to new greenhouse gas markets.
I then consider what the literature implies for two special features of international greenhouse gas trading, linking markets together and offsets.

**Lessons for designing enforcement strategies**

Perhaps the most difficult task of designing an effective market for greenhouse gas emissions is obtaining accurate emissions data. This task may actually be more difficult for greenhouse gas control than for other problems. The first difficulty is the sheer number of sources. The EU ETS covers more than 11,000 sources of CO$_2$ emissions. In addition, a comprehensive control policy would control emissions of several greenhouse gases. For example, the American Clean Energy and Security Act of 2009 would have controlled emissions CO$_2$, as well as methane, nitrous oxide, sulfur hexafluoride, perfluorocarbons, nitrogen trifluoride, and hydrofluorocarbons (US Congress 2009). Emissions of these gases can be more difficult to monitor than CO$_2$ emissions (McAllister 2011).

Emissions monitoring in the EU ETS is likely to be a reasonable approximation of the way monitoring will be done in future trading programs to mitigate climate change. While some sources may be required to (or opt to) employ continuous emissions monitoring systems in future programs, most programs will rely on estimates of emissions based on activity data and emissions factors. Self-monitoring and self-reporting of these data will continue to be important elements of enforcement. Moreover, future trading programs will likely allow sources the limited ability to bank and borrow permits. Allowing intertemporal permit trading has proven to enhance the cost-effectiveness of emissions markets and to help contain uncertain abatement costs.

Hence, future greenhouse gas markets, indeed future trading programs generally, are likely to involve imperfect monitoring, heavy reliance on self-reported data, and intertemporal
permit trading. The literature provides six general lessons for enforcing emissions markets in these environments.

1. There is no need for targeted enforcement based on firm characteristics in reasonably competitive trading programs. Knowledge of whether a firm buys or sells permits, of its abatement or production technology, or choices of inputs or levels of outputs cannot help a regulator make its enforcement resources more productive. This principle does not apply in cases of market power or significant transaction costs.

2. The key to enforcing a dynamic trading program with imperfect monitoring is motivating accurate self-reporting of emissions (or emissions proxies). Once the problem of getting accurate reporting is solved, motivating firms to hold sufficient permits to cover their emissions is relatively simple. On the other hand, permit compliance is not possible without reporting compliance.

3. High permit violation penalties, like those found in existing emission markets, have little deterrence value because of permit banking and the standard requirement that permit violations be offset with reductions in future permit allocations. Stiff sanctions for reporting violations can substitute for monitoring effort, while high permit violation penalties cannot.

4. Do not rely on noncompliance as a safety valve to control the risk of high permit prices. Instead, implementing an explicit price ceiling and floor and enforcing a trading program to
achieve full compliance can reduce expected enforcement costs and provide more efficient abatement cost containment.

5. Consider tying sanctions to observed permit prices. This form of sanction can absorb variation in permit prices, shielding enforcement effort and compliance outcomes from price changes.

6. Our first inclination should be to design enforcement strategies that achieve full compliance. This may seem to be an obvious point, but one could come away from reading much of the economic literature on environmental enforcement with the impression that it is normally efficient to design policies so that firms are noncompliant. Eliminating the variable costs of sanctioning is a powerful motivation for designing trading programs that achieve full compliance. Moreover, there are sufficient levers in the design of an emissions market to achieve full compliance, without incurring additional costs or altering the policy’s environmental outcome.

**Linking emissions markets**

International efforts to control greenhouse gas emissions now and into the extended future will likely consist of many independent national and regional policies, rather than a comprehensive global system. An important consideration, then, is whether and how to link independent emissions markets together. Although emissions markets can be linked together in several ways, it usually means that sources in one program can trade permits with sources in other programs. There are several potential benefits to linking emissions markets, but the most important is the
potential to reduce overall abatement costs. Discussions of some of the issues associated with linking greenhouse gas markets can be found in Kruger et al. (2007), Jaffe, Ranson and Stavins (2009), and Metcalf and Weisbach (2012).

An under-studied consequence of linking emissions markets is the likelihood that enforcement activities and the compliance performance of separate programs will impact each other when the programs are linked. Linking programs causes compliance incentives to change as permit prices adjust, and enforcement activities or compliance choices (or both) may change in response. Consequently, linking markets can mitigate or exacerbate enforcement and compliance problems.

To illustrate these issues, consider a simple model of two countries with domestic emissions markets for CO₂ emissions that are unlinked at first. Imagine in country A that enforcement is sufficient to induce full compliance by all covered sources, but that country B has a more difficult time enforcing its cap so it experiences significant noncompliance. For simplicity, consider only permit violations. Assume that penalties are not tied to permit prices, that the expected marginal penalties in both countries are increasing, and that total permit supplies remain constant after the programs are linked together.⁶

Suppose at first that the price of CO₂ permits is higher in country A than in country B. Differences in permit prices for unlinked systems can be generated by differences in permit supplies, aggregate marginal abatement costs, and enforcement effectiveness. In fact, permit prices in country B could be lower, in part, because of weak enforcement. If the countries link their programs together, sources in A will purchase permits from sources in B until the prices in the two countries are equalized. The price decrease in country A produces greater emissions in

---

⁶ However, Helm (2003) argues that countries may choose their national caps in anticipation of linking with other programs, and shows how this can lead them to choose lower caps.
this country, while its sources remain compliant. In contrast, the increased price in country $B$
causes greater violations but lower emissions. It is straightforward to show that this reduction in
emissions is more than offset by the increase in emissions from country $A$, so aggregate
emissions and violations from the two countries increase as a result of linking.

It is also possible that linking can mitigate compliance problems under some
circumstances. Suppose that the same conditions apply except that the price of permits is higher
in country $B$ than in $A$ when the programs are not linked. Perhaps the high permit price is one
reason that country $B$ authorities find it difficult to maintain full compliance. Linking the
programs together would result in country $B$ sources purchasing permits from country $A$ sources,
increasing the permit price in $A$ and reducing the permit price in $B$. The price increase in $A$ would
lead its sources to reduce their emissions, but could lead to noncompliance that did not exist
before linking. On the other hand, violations in $B$ would be lower even though sources there
would increase their emissions. In the end, linking in this case could result in lower aggregate
emissions if violations in country $B$ fall by more than they increase in country $A$.

The interconnection of enforcement and compliance performance continues after
programs are linked together. In fact, price risk can be transmitted across programs, which has
compliance and enforcement consequences. For example, suppose that the programs in countries
$A$ and $B$ are linked together, and for some unforeseen reason marginal abatement costs in country
$A$ are higher than expected. This will result in a higher permit price in both countries, increasing
the noncompliance incentive and placing additional strain on enforcement resources in both
countries. Tying sanctions directly to prevailing permit prices may help mitigate the price effects
on compliance and enforcement that are transmitted across linked markets.
All this suggests that careful consideration be given to the enforcement consequences of linking markets, both in the implementation phase as prices and compliance incentives adjust to the initial inking, and in administering already-linked markets as shocks in one market can change prices and compliance incentives across markets.

Offsets

Many existing and proposed emissions trading programs allow covered sources to purchase abatement from outside sources. Offset credits can be generated in many ways, including reducing greenhouse gas emissions from industrial facilities, reducing methane emissions from coal mines, wastewater treatment plants, and landfills, and sequestering carbon in forests and soils. The main expected benefit of allowing offset credits is that the costs of an emissions market can be reduced if there is a significant supply of low-cost offsets available. Active offset markets may yield additional benefits by producing incentives for technological advance and by promoting the development of institutional capacity to control greenhouse gas emissions in other countries (Sigman and Chang 2011).

However, there are serious enforcement concerns with making certain that offsets represent real emissions reductions. The number and variety of potential offsets makes monitoring their emissions reductions more complex than monitoring emissions from large point sources in the typical emissions market. Offset credits may come from a large number of smaller concerns and measuring carbon sequestered in forests and soils may be more difficult than measuring greenhouse gas emissions from a large industrial source. Moreover, some projects must be monitored on a permanent basis, because the emissions reductions can be reversed in the
future. For example, carbon sequestered in a forest can be released if the forest is burned to clear the land for agriculture at a later date.

In addition, monitoring activities are not simply limited to the performance of an offset project. Typically, offset projects must be additional, meaning that the emissions reduction would not have occurred without the project. Establishing the additionality of an offset project requires gathering information to establish a business-as-usual baseline. A related problem is that of leakage, whereby emissions reductions simply produce increased emissions elsewhere. For instance, a program to preserve forested land to sequester carbon may motivate increased timber harvests in another area. Thus, assessing the performance of an offset project may require additional monitoring to determine whether the controlled activity moved elsewhere.

Monitoring offset performance and sanctioning nonperformance is further complicated by the fact that the trade of offsets often involves different legal jurisdictions (Bushnell 2012). The regulatory authorities of an emissions market and those of another jurisdiction that produces offsets must agree on and commit to procedures for monitoring the performance of offsets and levying sanctions in case they do not produce the required emissions reductions. Additional monitoring and enforcement difficulties come from differences in the regulatory and legal capacities of regulators in the offset trade. Large supplies of offset credits from reducing emissions from deforestation and degradation (REDD activities) are potentially available from developing countries. However, some of these countries lack the institutional capacity for environmental monitoring and enforcement, as well as the control of corruption, that would give their offsets credibility (Murray, Lubowski and Sohngen 2009; Deveny et al. 2009).

These enforcement challenges will likely lead to compliance problems in offset markets, as some credits are traded that do not represent actual emissions reductions. However, this does
not necessarily imply that offset provisions are counterproductive in terms of controlling aggregate emissions or producing more efficient trading programs. In fact, Sigman and Chang (2011) show theoretically how offset provisions can produce higher abatement and lower emissions even if offsets involve higher enforcement costs and significant noncompliance. Their argument is straightforward. Suppose that enforcement of a particular emission market is not sufficient to induce full compliance. Assume further that sanctions in this program do not vary with the permit price. In this case, allowing purchases of offset credits outside the program will reduce the price of permits, which in turn will lead to reduced noncompliance. Some portion of these offset credits may not be legitimate. However, violations and emissions will be lower if the reduction in violations of sources in the regular emissions market is larger than the violations in the offset market. In this case, the offset provision unambiguously enhances the efficiency of the program despite higher enforcement costs and noncompliance in the offset market.

Efficiency can be improved even if offset provisions lead to higher emissions. If sources under a trading program are fully compliant and they are allowed to purchase offset credits, some of which are illegitimate, then aggregate emission will increase. However, aggregate abatement costs in the program will be lower, and the lower permit price that comes from trading offsets could lead to a reduction in the costs of enforcing the program. The offset provision could enhance efficiency if the reduction in abatement and enforcement costs is greater than the costs of the legitimate offsets, their enforcement costs, and the environmental damage associated with higher aggregate emissions.

I do not wish to argue that allowing offset credits always improves emissions markets. Enforcement and compliance problems associated with offsets could easily lead to worse environmental and economic outcomes. The point is that these problems are not inevitable.
Careful estimates of the compliance and cost consequences of adding offset provisions to emissions markets are necessary to determine whether these provisions are worthwhile.

**Conclusion**

Experience and scholarship has revealed much about the compliance and enforcement challenges that are unique to emissions markets. In particular, the economic literature has provided important insights into the nature of sources’ compliance incentives and the effective and efficient design of enforcement strategies. The focus of this paper has been on highlighting these insights.

However, the enforcement challenges of emissions trading have perhaps become more difficult as markets are developed to control greenhouse gas emissions. These challenges include the monitoring and enforcement difficulties associated with controlling multiple pollutants from a wide variety of sources, international trade of greenhouse gas emissions among independent emissions markets, and the management of offsets. While existing economic models and empirical tests can provide important lessons for the enforcement of these newer programs—and we have examined many of these lessons—the literature must continue to develop as new regulatory innovations emerge to confront climate change and other environmental problems.

**Acknowledgements:** I wish to thank Carlos Chavez, James Murphy, and John Spraggon for many of the ideas in this review that originated in our collaborations. This research has been supported by a grant from the U.S. Environmental Protection Agency's National Center for Environmental Research (NCER) Science to Achieve Results (STAR) program. Although the research described in the article has been funded wholly or in part by the U.S. EPA through grant
number RD-83367201-0, it has not been subjected to the EPA’s peer and policy review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.
References


Murray, Brian, Ruben Lubowski and Brent Sohngen. 2009. Including international forest carbon incentives in climate policy: understanding the economics. Duke University Nicholas Institute for Environmental Policy Solutions, Report NI R 09-03.


