When external effects are important, markets will be inefficient, and economists have considered several broad classes of economic instruments to correct these inefficiencies. However, the standard economic analysis has tended to take the region, and the government, as a given; that is, this work has neglected important distinctions and interactions between the geographic scope of different pollutants, the enforcement authority of various levels of government, and the fiscal responsibilities of the various levels of government. It typically ignores the possibility that the externality may be created and addressed by local governments, and it does not consider the implications of decentralization for the design of economic instruments targeted at environmental problems. This paper examines the implications of decentralization for the design of corrective policies; that is, how does one design economic instruments in a decentralized fiscal system in which externalities exist at the local level and in which subnational governments have the power to provide local public services and to choose tax instruments that can both finance these expenditures and correct the market failures of externalities?

Keywords: market failure, environmental federalism, externalities, fiscal decentralization, subsidiarity principle, economic instruments

JEL: H2, H4, H7, Q5
DESIGNING ECONOMIC INSTRUMENTS FOR THE ENVIRONMENT IN A DECENTRALIZED FISCAL SYSTEM

James Alm and H. Spencer Banzhaf*

Abstract

When external effects are important, markets will be inefficient, and economists have considered several broad classes of economic instruments to correct these inefficiencies. However, the standard economic analysis has tended to take the region, and the government, as a given; that is, this work has neglected important distinctions and interactions between the geographic scope of different pollutants, the enforcement authority of various levels of government, and the fiscal responsibilities of the various levels of government. It typically ignores the possibility that the externality may be created and addressed by local governments, and it does not consider the implications of decentralization for the design of economic instruments targeted at environmental problems. This paper examines the implications of decentralization for the design of corrective policies; that is, how does one design economic instruments in a decentralized fiscal system in which externalities exist at the local level and in which subnational governments have the power to provide local public services and to choose tax instruments that can both finance these expenditures and correct the market failures of externalities?

Keywords: Market failure; Environmental federalism; Externalities; Fiscal decentralization; Subsidiarity principle; Economic instruments

JEL Classification Codes: H2, H4, H7, Q5

* Tulane University and Georgia State University. We are grateful to the editor and to an anonymous referee for many helpful comments. We are also grateful to the Municipal Finance and the Environmental Economics Thematic Groups of the Sustainable Development Vice-Presidency of The World Bank for financial support, and especially to Patricia Annez and Jan Bojo for many helpful comments. The views here are those of the authors, and do not represent the policy of The World Bank or its Board of Directors. Please address all correspondence to: James Alm, Department of Economics, Tulane University, 6823 St. Charles Avenue, 208 Tilton Hall, New Orleans, LA 70118 USA (phone +1 504 862 8344; fax +1 504 865 5869; email jalm@tulane.edu).
1. Introduction

In many settings, markets can be a highly effective way to organize an economy, efficiently making the most of economic resources. In such a system, agents compare private benefits and private costs, ignoring external costs of their actions such as pollution. However, when these external effects are important, when property rights are ill defined, or when the impediments to contracts are high, markets will not be efficient.

To correct such deficiencies, economists have considered three broad classes of economic instruments (EIs) (Cointreau and Hornig, 2003). First, if pollution is measurable, the government may set a “Pigouvian tax” on pollution equal to the marginal external cost of environmental damage, thereby providing the correct incentive to private agents. Second, the government may set a “cap” on pollution, allowing trades or offsets so that the total amount of pollution does not exceed the cap, but with the market determining the specific agents who do the polluting. Third, if pollution is not measurable, the government may impose “command and control” regulations, for example, by requiring that a factory use a particular pollution abatement technology.

There is now a vast literature comparing and contrasting these approaches to controlling pollution (Sterner, 2003). However, the standard analysis in this literature has tended to take the region, and the government, as a given. Consequently, it has tended to neglect important distinctions between the geographic scope of different pollutants, the enforcement authority of various levels of government, the fiscal authority of various levels of government, and the interactions among these factors. For example:

1. Externalities generated in a particular local, or subnational, area may be confined to the local area or may spill over to other jurisdictions.

2. Local governments may be better informed about how best to regulate or enforce pollution control within their jurisdiction, but they may not consider the effects of their actions on other jurisdictions.
3. The existence of locally-generated waste emissions affects the appropriate assignment of both expenditure and tax responsibilities among levels of government. The Pigouvian corrective tax (or other economic instruments) may be chosen by the local government in various ways, either making its decision alone and ignoring the actions of other local governments, or making its decision while recognizing that its actions may elicit responses by other local governments as they respond in a strategic manner.

The standard analysis therefore focuses mainly upon an aggregate, or national, perspective and typically ignores the possibility that the externality may be created in, and addressed by, local governments. Put differently, the standard analysis does not consider fully, if at all, the implications of decentralization for the design of economic instruments targeted at environmental problems. (For purposes of discussion, we consider a “subnational” government as any government below the national level, thereby including state or provincial, county, municipal, village, or special district governments.)

This paper examines the implications of decentralization for the design of economic instruments; that is, how does one design EIs in a decentralized fiscal system in which externalities exist at the local level and in which subnational governments have the power both to provide local public goods and services and to choose tax instruments that can finance these expenditures and also correct the market failures of externalities?

In section 2, we review the economic theory on EIs and consider our experience with them to date. From both the theory and evidence, we draw several rules of thumb about which EIs are most likely to be successful in various contexts. In section 3, we consider a related literature on public finance in a federal context. Well-established lessons about taxation and public goods provide insights into the problem of pollution control. In section 4, we synthesize these literatures, drawing guidelines for the control of pollution under decentralization. A final section 5 summarizes our main conclusions.
Externalities and Market Failure: The Standard Approach

The standard analysis of an externality starts with welfare economics and the conditions for Pareto efficiency, defined as a state in which no one can be made better off without adversely affecting someone else (Cornes and Sandler, 1986; Tresch, 2002). A competitive market will operate where marginal private benefits to consumers equal marginal private costs to producers. Under some conditions, notably the absence of external effects in either production or consumption, the market will also achieve Pareto efficiency, producing a level of output at which marginal social benefits equal marginal social costs.

However, in the presence of external effects like pollution, private benefits and costs diverge from social benefits and costs. In particular, when the production of a product also generates harmful emissions, there will be external, or spillover, costs from production. Because a competitive market will produce where private benefits and costs are equal, and will ignore these external costs, a competitive market will no longer achieve Pareto efficiency, producing both too much of the product that generates the pollution and doing so in a way that is too “dirty”.

A comparable analysis focuses more directly on the emissions that characterize the pollution-generating activity, and is useful for subsequent discussion of decentralization. Suppose that the utility of a (representative) consumer depends upon his or her consumption both of a private good and of the level of pollution. Pollution is generated by emissions, or waste discharges, that result from the production of the private good, and production of the private good depends positively upon the use of traditional inputs (e.g., capital and labor), negatively upon the level of pollution, and positively upon the quantity of emissions. Emissions are
therefore treated as a factor of production like capital and labor, so that production of the private good can be increased by the use of more capital or more labor but also by the discharge of more waste because more waste discharges do not require the use of other inputs to limit their use.

In this framework, it is straightforward to demonstrate that a competitive market will generate a level of emissions where the marginal private benefits of emissions are driven to zero, because emissions are effectively a free input whose use does not require the payment of any factor payments comparable to, say, wages paid for labor or interest paid for capital. Accordingly, the market outcome will lead to an inefficiently large amount of emissions, because private firms that generate the emissions will not consider the external costs of emissions on consumers (who are hurt by more pollution) or on other firms (whose production is negatively affected by more pollution). See Cornes and Sandler (1986), Hahn (1989), Newberry (1990), and Cropper and Oates (1992) for useful discussions of much of this literature.

However, as noted earlier, there are several corrective policies that a government can pursue. The simplest corrective policy is a Pigouvian tax, equal to the marginal external costs of the pollution at the efficient level of output (Baumol, 1972; Barthold, 1994; Bovenberg and de Mooij, 1994; Bovenberg and Goulder, 1996, 2002). The imposition of such a tax brings private costs into line with social costs and thereby achieves the efficient level of production of the good. Polluters who have the lowest cost of abatement will reduce pollution the most, and polluters who have the highest cost will do less abating. In addition to such static efficiency, Pigouvian taxes have the dynamic property of creating an incentive (namely, a reduction in tax payments) for entrepreneurs to develop more efficient ways to reduce pollution. Although they are simple in theory, implementing such taxes in practice raises challenges related to the
measurement of the marginal external costs and the monitoring of the pollution levels subject to taxation.

Closely related to a tax on pollution is a subsidy on pollution abatement or on cleanup. In this EI, firms face the same marginal incentive as a tax, but the incentive is in the form of a “carrot” (e.g., a subsidy) rather than a “stick” (e.g., a tax). This has the advantage of making the instrument more politically palatable, but it also has the disadvantage of encouraging more activity in a dirty industry, which is counterproductive (Baumol and Oates, 1988). Nevertheless, if polluters cannot be identified or if taxes can be avoided by, say, “midnight dumping” (Fullerton and Kinnaman, 1996), there may be no alternative.

An alternative EI is a cap-and-trade system, in which a limit is set on the total quantity of pollution allowed. This aggregate cap is then allocated to individual polluters through a permitting process, and may either be auctioned (to raise revenue) or distributed without charge. The holders of the pollution permits may in turn trade them to another entity for a mutually-agreed upon price. Although seemingly different on the surface, such a system is quite similar to a tax: just as under a Pigouvian tax they must pay a tax for each unit of pollution, in the cap-and-trade system polluters must purchase a permit (or must forego the opportunity of selling one). Thus, a cap-and-trade system is formally equivalent to a Pigouvian tax: if a cap allowing pollution level $Q$ leads to permits trading at a price $P$, then a tax set at $P$ will lead to a level of pollution $Q$. However, when future economic conditions are uncertain, they differ insofar as a properly implemented cap-and-trade system always guarantees the level of pollution ($Q$) but not the cost ($P$), whereas the tax system guarantees that the marginal cost of cleaning up pollution will be $P$ but does not guarantee that the level of reduction will be $Q$ (Weitzman, 1974).
The foregoing EIs all work through economic incentives. There are also other types of corrective policies, including ordinary regulation, frequently called “command and control” (CAC). Depending on the nature of the regulation, CAC policies may often be the easiest to administer. For this reason, they are by far the most common approach to pollution control. However, they do not have the same efficiency properties as EIs (Bergstrom, 1976; Hahn, 1990; Hazilla and Kopp, 1990). For example, they typically impose a one-size-fits-all mandate on all polluters, making them especially burdensome when there is a great deal of variability in the costs for different polluters to clean up. They also typically do not provide incentives for polluters to find better ways to reduce pollution.

As summarized by Cointreau and Hornig (2003), these EIs can be broadly classified into three categories: revenue-generating instruments, or those EIs that produce revenue for governments (e.g., charges, taxes, reductions in subsidies, or auctioned permits); revenue-providing instruments, or those EIs that allow producers and service providers to receive income from governments (e.g., fiscal incentives, development rights, or charge/tax reductions); and non-revenue instruments, such as deposit-refund schemes and grandfathered permits. Traditional CAC regulation is also revenue-neutral.

**Economic Instruments in Context: Legal, Institutional, and Cultural Considerations**

This standard analysis of economic instruments for pollution control is an elegant application of economic theory, and it provides powerful insights into the advantages of EIs over traditional command-and-control instruments. However, like all models, it carries assumptions that must be carefully identified and evaluated before actually implementing any EI, especially in a decentralized fiscal system. Adopting an instrument that is cost-effective only under particular circumstances may be quite ineffective if those circumstances do not hold.
There are, we believe, five key requirements for EI mechanisms to be effective.

First, participants (i.e., individuals or firms) must be familiar with market prices or, in the context of tradable permits, with market transactions. This requirement is probably the easiest to meet, especially when applied to firms. However, households in some cultures may be more familiar with bartering and haggling, and may be less familiar with a posted price. Even if they are familiar with posted prices in some contexts, if their behavior in, say, water use is guided more by tradition or habit, rather than by constant re-optimization, they may not be responsive to taxes or fees set at politically acceptable levels; even so, the standard theory can accommodate this point by simply describing it as a case of an extremely inelastic demand.

Second, in the case of firms, another key requirement for EI mechanisms to be effective is that decision makers must maximize profits; at a minimum, profit must be one important objective. Even in highly developed countries such as the U.S. and those in Europe, this point cannot be taken for granted. Principle-agent problems may arise when actual decision makers do not have the interests of the firm’s owners. A plant manager, for example, may care more about making his work-day run smoothly than maximizing the profits of shareholders whom he has never met. In developing countries, especially in communist or former-communist countries, managers may have an administrative mindset in which they focus on meeting output goals rather than profit goals. The potential for this problem is compounded by the presence of “moral hazard”, in which counterproductive actions are taken by agents in response to incentives created by government policies. For example, even if managers are keenly watching the bottom line, they may also be shrewdly aware of the fact that costs incurred from paying pollution taxes or fees, or from purchasing pollution permits, can be recovered from governments or from parent companies. Such “soft budgets” (Kornai, 1979) may show up in the form of subsidies, lowered
taxes, or more favorably administered prices. By the same token, subsidized pollution abatement may trigger reductions in these side-payments.

Western experience shows that such results can be quite subtle. For example, consider the case of U.S. electric utilities. This industry operates under a cap-and-trade system for sulfur dioxide (SO$_2$) pollution, in which the industry receives a certain number of free permits to cover a limited quantity of emissions but which can be bought or sold by firms at prevailing market prices. When the electricity sector operates competitively, this cap-and-trade system will tend to increase electricity prices because generating more electricity requires either purchasing more permits to cover the associated pollution or making increased expenditures on abatement. These higher prices in turn provide an incentive for consumers to conserve electricity, in addition to the incentive for firms to operate more cleanly. However, the electric utility industry has traditionally operated as a regulated monopoly, charging an administered price for electricity that reflects a fixed mark-up over average costs. Despite a wave of de-regulation in the 1990s, much of the industry continues to operate under this arrangement. Parry (2005) shows that, in this context, increased marginal costs can be offset by the decline in average costs due to the lump-sum transfer of the emissions quota. In such a case, although electricity will be more cleanly produced on a per-unit basis, prices will not increase, and there will be no signal to consumers to conserve energy.

In developing countries and, again, especially in communist or former-communist countries, such effects may be quite important. Söderholm (2001) emphasizes the role that administered prices, centrally granted investments, and output targets still hold in Russia, all of which undermine the effectiveness of pollution charges. Bell (2005) argues that the same issues are relevant for China, where much of the industry continues to be directed by the state.
A third requirement for EI mechanisms to function properly is the monitoring of pollution. Obviously, if emissions fees are to be collected, or if tradable permits are to be required to be held to cover emissions, then those emissions must be observed in a cost-effective manner. For this reason, non-point sources of pollution (such as agricultural run-off or automobile tailpipe emissions) are rarely if ever covered by EI mechanisms. Rather, even developed countries that embrace EI mechanisms in other contexts use command-and-control regulation such as best-management practices for agriculture and catalytic converters for cars. They may also tax proxies for pollution such as gasoline. Municipal solid waste provides another intriguing example. Although some cities have experimented with charges by weight or volume, such payments can be easily avoided with “midnight dumping”. In these cases, a deposit-refund system, or revenue-providing instruments such as a subsidy to recycling, may be preferable to a simple tax because actors have the incentive to prove their compliance, so as to collect their refund or subsidy.

Relatedly, EI mechanisms are more appropriately used where they can be observed at a manageable number of discrete points (e.g., filling stations, smoke stacks, or water pipes) and where those points cannot be easily bypassed via leaks in the system. Even where more likely to be successful, the continuous monitoring systems that must be installed to monitor and record emissions can be expensive and sophisticated equipment. The expense of such systems is sometimes given as a reason that EI mechanisms are not feasible even for large point sources in poor countries. However, it can also be seen as a concrete opportunity for donors to assist pollution-control efforts.

Fourth, enforcement of any violations must be consistent and sufficient to discourage cheating. If polluters know that they will not pay a penalty even if detected, then they will not
have the incentive to reduce pollution. Weak laws, corrupt prosecutors or courts, and a lack or inability to bring legal action on behalf of the public all have the potential to undermine the enforcement of environmental provisions (Bell, 2005). In the case of revenue-generating instruments, such enforcement might be expected to be similar to the enforcement of tax laws. So long as pollution fees are covered by the same branch of law, performance with public finances may indicate the success of future performance with EI mechanisms. There is in fact a large literature on “optimal enforcement” schemes (Kwerel, 1977; Segerson, 1988; Kritikos, 2004).

Fifth and finally, it is important for any environmental policy instrument to be accepted by the participants involved. Not all instruments are familiar or acceptable in all cultures. It took many years of discussion before a cap-and-trade system could be tried in the U.S., and more years of experience with early programs before it became widely accepted. Notwithstanding recent trading in carbon, these systems still are not used widely in Europe. On the other hand, pollution taxes, used widely in Europe, are used infrequently in the U.S. One can speculate about the reasons: Americans are more willing to commoditize anything and more opposed to taxes, Europeans are more interested in making polluters pay, Americans are more focused on achieving a specific quantity target, and so on (Harrington, Morgenstern, and Sterner, 2004). Whatever the reason, the important point is that not all instruments will go over equally well in all cultures, even if the economic conditions are the same.

We have focused here on requirements for EIs to be fully effective. However, we do not want to leave the impression that, wherever any of these requirements are not fully met, EIs will be a complete failure, or that CAC or other non-revenue and regulatory instruments will be preferable. Each requirement is one of degree, and must be assessed qualitatively. Moreover,
many of these considerations are important for CAC instruments as well. For example, a traditional CAC regulation requiring the installation of some specific pollution-reducing equipment too will fail if violations of the regulation are not detected and enforced. Policies must be evaluated according to these criteria on a case-by-case basis, and conclusions will no doubt differ for different societies and different pollution problems.

**Some Experience with Economic Instruments**

Before implementing any sophisticated market strategies for reducing pollution, it is essential to examine the experience we have already had with EIs, most of which has occurred in developed countries. Such experience is extensive, and covers an array of instruments for all types of pollutants. In fact, multiple instruments are often used for a single pollution problem in a single country. A recent set of six pairs of case studies, examining the treatment of a particular type of pollution in the U.S. and Europe, found in 11 of 12 cases that a mixture of several tools was used, including both CAC and EIs (Harrington, Morgenstern, and Sterner, 2004).

The U.S. acid rain program, which instituted tradable permits for SO$_2$ emissions on electric utilities, provides an example. All plants, new and old, fall under the cap on total emissions, perhaps the most important EI precedent (what Stavins (1998) has called “the grand experiment”). At the same time, new plants must satisfy new source performance standards, which require meeting a maximum rate of emissions per unit electricity, a rate usually set according to the “best available control technology”. These performance standards were a legacy of older regulatory approaches kept because of distrust of the EI instrument, and can now be justified as a way to guarantee that trading patterns in pollution permits do not create “hot spots” in a particular area. Unfortunately, they also provide an incentive to keep older, dirtier power plants (which by law are not required to meet the standard) in operation longer. To offset
this perverse effect, a costly process known as “new source review” examines major investments 
at older plants to determine whether they effectively made the plant a new source for regulatory 
purposes.

What can we learn from these experiences? We draw four key lessons; see Harrington, 

First, as textbook theory predicts, EIs are more efficient than CAC regulations. For 
example, in the U.S. SO₂ cap-and-trade program, permit prices have traded substantially lower 
than predicted, suggesting that cost-saving methods of reducing pollution emerged in response to 
the economic incentives. Carlson et al. (2000) estimate that the cost savings from trading are on 
the order of $800 million per year, or 43 percent from a uniform emissions rate. Ellerman et al. 
(2000) similarly estimate the savings to be about 55 percent. In contrast to this success, 
Sweden’s outright ban of trichloroethylene, a toxic solvent used in metal degreasing, met with 
failure and was repealed. Protests came from a small number of firms experiencing especially 
high costs (Sterner, 2004). As noted previously, such cost heterogeneity is a prime example 
where EIs are preferable to one-size-fits-all regulations.

Second, because the gains from trade are limited by the extent of the market, EIs work 
best when they pool together a large number of firms, and they have their greatest advantage 
when those firms have different costs of compliance. The Swedish ban on trichloroethylene 
illustrates this point. The policy was effectively derailed because only a small number of firms 
faced especially high costs of compliance, and such firms would have stood to gain from the 
flexibility of an EI policy, perhaps by buying pollution credits. Similarly, in the United States, 
an early experiment with cap-and-trade along a single river valley (Wisconsin's Fox River) met
with failure, as there were too few firms involved in the program. In contrast, the larger-scale SO₂ trading program has been successful.

Third, EIs can be more efficient dynamically than CAC instruments. As the U.S. phased out lead in gasoline in the 1980s, it allowed trading to aid small and struggling refineries. It also allowed emissions banking, in which emissions could be reduced by a greater amount in the present to allow emissions above a (tighter) cap at a later date. Newell and Rogers (2004) estimate that the banking provision alone saved more than $225 million, with the total program saving hundreds of millions more. Even without banking, EIs can provide incentives to identify new ways to reduce pollution more cheaply. Such dynamic effects are most likely to be important for highly technical problems such as gasoline refining (where substitutes were found for lead) and air pollution reductions. In the latter case, although there have been few new technologies invented, experimentation with processes (e.g., fuel blending) has led to important discoveries that have reduced costs. This consideration may be less important for situations that require basic actions, or actions that are likely to be part of any cost-effective solution for the foreseeable future. Examples might include basic sanitation or buffers between livestock and water sources.

Fourth, administrative, monitoring, and enforcement costs can differ on a case-by-case basis. For example, as suggested by the Swedish trichloroethylene experience, cost heterogeneity can cause administrative problems for CAC regulations. In contrast, pollution fees and taxes or permits allow more pollution at higher marginal-cost facilities. Many EIs have been implemented in Europe without any reference to firm-level costs (Harrington, Morgenstern, and Sterner, 2004). Generally, then, monitoring is more costly and demanding for EIs. Again, however, some aspects of the problem can be subtle. For example, in the U.S. lead phase-out,
pollution standards were set in terms of lead content (i.e., per unit gasoline). As it turns out, monitoring lead was fairly straightforward because it could be checked against sales figures from suppliers. Gasoline volume (the denominator) was harder to monitor, so a straight cap on lead use rather than on lead content would have been easier to enforce (Newell and Rogers, 2004).

One advantage of incentive-based instruments is that they are often defined objectively, in terms of physical outputs. In contrast, some CAC regulations, such as the use of a best available control technology, are quite subjective, and require substantial executive rule-making, which in turn may be challenged in court. The U.S. experience with such regulations has involved substantial legal proceedings. In many cases, non-governmental organizations (NGOs) have stepped in to require compliance to fill the void left by executive agencies (Bell, 2005). However, European nations do not seem to have had these difficulties (Harrington, Morgenstern, and Sterner, 2004).

Implications for the Choice of Pollution-control Mechanisms: The Special Case of Developing Countries

As emphasized previously, evaluation of economic instruments must be made on a case-by-case basis. However, the above principles and experiences provide several rules of thumb that provide a starting point for thinking about pollution controls in developing countries. After considering the fiscal federalism literature in the following section, we return to this question in the context of decentralization.

There are several main implications. First, in cases where monitoring and enforcement is weak, more objectively measured standards and behaviors should be targeted. Emissions standards, taxes and fees, or CAC regulations requiring very specific actions may all be preferable to standards open to more interpretation. If requirements such as a best available
control technology or an adequate margin of safety create confusion, years of administrative review, and years more of legal challenges even in an open country such as the U.S., they may face even more difficulties in developing countries. Emissions standards or emissions fees, while in some ways quite sophisticated, are also quite objective in their measurement of a physical unit. Here, donors can help by providing or financing expensive monitoring equipment and assisting with their use. In other cases, where enforcement and the rule of law are strong but resources are not available for emissions monitors, trading based on proxies, or modeled emissions, may be a feasible alternative. Montero (2005) reports success with such a program in Santiago, Chile, where tradable permits are required to be held, not for actual pollution but for a formula based on the technologies employed at each facility.

Revenue-providing instruments may also be useful. Far from having an incentive to cheat, in these cases polluters have an incentive to prove their compliance in order to collect a payment. Subsidies thus overcome a monitoring problem, but this comes at an efficiency cost. While they provide the correct marginal incentive for polluters to produce their product in a cleaner way, they subsidize the polluting activity, so that more firms may enter the polluting industry (Baumol and Oates, 1988). On net, subsidies thus may not be effective, especially where firms can easily enter the industry and where abatement costs represent a large fraction of profits. However, they may be quite effective if used in combination with another EI. For example, a deposit-refund is essentially a two-part instrument: a subsidy that encourages recycling combined with an offsetting tax (or the deposit) that eliminates the accompanying perverse incentive to purchase more of the raw material (Fullerton and Kinnaman, 1995). In the same way, a tax on dirty cars and/or gasoline usage can be used in combination with a subsidy on pollution control equipment to mimic a pollution tax (Fullerton and West, 2002).
Second, where the profit motive is weak, rules may be preferable to those mechanisms that rely more upon economic incentives. Emission standards and technology-based standards are consistent with the administrative mindset of firms operating in such contexts, or firms that might not respond to fees or subsidies. As noted previously, such dynamics can arise for firms operating in the context of soft budget constraints, as in the transition countries. Consider another example provided by China. With assistance from the Asian Development Bank, China has experimented with an SO$_2$ pollution trading system in its Taiyuan province (Morgenstern et al., 2005). However, so far the experiment has been a disappointment, with few trades (Predd, 2005).

Table 1 summarizes these recommendations. Where the rule of law is strong, polluting activities are observable, and profit motives are strong, the full power of EIs can be unleashed to achieve static and dynamic cost-savings. Where the rule of law is weaker, polluting activities are hard to observe, but profit motives remain strong, subsidies (perhaps in combination with other instruments) may be a better approach. Where the rule of law is strong but the profit motive is weak, traditional CAC regulation may be the best option. Finally, where the rule of law and profit motives are weak, environmental improvements will be particularly challenging.

<table>
<thead>
<tr>
<th>Rule of Law; Observability of Activity</th>
<th>Profit Motive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak/Low</td>
<td>Weak/Strong</td>
</tr>
<tr>
<td>Strong/High</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>Subsidies/Two-part Instruments</td>
</tr>
<tr>
<td></td>
<td>CAC</td>
</tr>
<tr>
<td></td>
<td>Pollution Fees/Permits</td>
</tr>
</tbody>
</table>

As a third and final recommendation, the chosen instrument should be acceptable to all relevant parties. It took some time for permits to be accepted in the U.S., but this acceptance did not come easily. The provisions allowing trading were scrutinized and tested by skeptics over
several years in the 1980s; industry was required to install continuous emissions monitors; permits were assigned serial numbers so they could be tracked and not easily counterfeited; and transactions were recorded. Only after these steps were taken was a consensus reached (Bell, 2005). The lesson we draw from this experience is that, if outsiders are beginning to plan an environmental policy by thinking about the best instrument, then they have already missed a step. The first step, not to be missed, is to think about the right process for choosing an instrument, one that includes local stakeholders.

3. Decentralization: The Assignment of Fiscal Responsibilities

In addition to the factors discussed above, the appropriate choice of instrument must reflect the degree of decentralization in a nation's environmental and fiscal policies. To understand the importance of this factor, we draw on the larger literature on fiscal decentralization, which has long addressed the question of the appropriate assignment of fiscal functions to various levels of government. See Oates (1972, 1993, 1999), Bahl and Linn (1992), Breton (1998), and Wellisch (2000) for useful discussions of much of this literature.

In the following subsections, we separately consider in detail each side of the fiscal coin, determining expenditures on public goods and raising revenues to finance those expenditures. However, because these two sides must be considered in light of the whole fiscal system, we provide a brief overview of some of the more important factors at the outset.

First, when tastes, incomes, and needs differ across regions, local governments will be in the best position to determine the expenditure priorities of its citizens.

Second, when local governments are in this position, they should bear the costs of financing those expenditures. Only then will they balance the benefits of public goods with the
costs; if the marginal cost of local public goods is subsidized by national governments, local governments will be apt to overspend.

Third, and mitigating the second point, we must consider the degree of spillovers across jurisdictions. If a public good provides benefits not only locally but across jurisdictions, a local jurisdiction may discount some of those benefits and under provide the good. In this case, a higher level of government may be in a better position to provide it; alternatively, it may need to subsidize local governments’ expenditures. Another potentially mitigating factor here is the cost of providing the public good. When there are economies of scale, then services can be provided more efficiently at larger scales than in a single local jurisdiction. In this case, it may be more appropriate for a higher level of government to provide the public good.

Fourth, and also mitigating the second point, revenues should be raised by taxing the most immobile tax bases. If a government tries to tax a mobile factor, the factor can easily avoid the tax by moving outside the relevant jurisdiction, thereby leading to a loss in revenues and causing distortions in the economy. Generally speaking, because it is easier for households, firms, and economic activities to move within a nation than across nations, factors are less mobile from the perspective of a central government than from that of a local government. This may be one reason for central governments to subsidize local governments' expenditures.

Fifth and finally, local governments may interact strategically, competing to attract and/or hold a larger share of mobile tax bases. This phenomenon has been characterized as a “race to the bottom”, as it suggests poorer quality of public services as local governments collectively cut tax rates.

With these initial insights in mind, we discuss in more detail expenditure responsibilities, followed by tax responsibilities.
The Assignment of Expenditure Responsibilities

The standard normative discussion of this question follows Musgrave (1959) by dividing the broad functions of expenditure responsibilities of government into three branches: the stabilization, distribution, and allocation roles. The stabilization function refers to smoothing business cycles, reducing inflation and unemployment, encouraging economic growth, and obtaining other related macroeconomic objectives. If the economic criteria for assignment are efficiency and equity, then it is typically concluded that the stabilization function should largely be performed by the central government because the mobility of resources makes it unlikely that an effective stabilization policy can be pursued by a lower level of government (Gramlich, 1987). In addition, local governments have limited powers to borrow or to issue money.

Redistribution also properly belongs to the central government. Subnational attempts to redistribute income are likely to be thwarted by the mobility of high-income individuals and of capital, and the attempts to redistribute income will create distortions and inefficiencies in geographic location (Oates, 1972). Moreover, the unequal and possibly inadequate fiscal capacities of local governments also make centralization desirable on equity grounds.

However, the allocation function, or the decision to provide local government services, should often be performed by local levels of government. These governments can adapt service levels more closely to the preferences of their citizens, thereby making available to individuals a wider range of fiscal choices than could be provided by uniform central government provision. This is the well-known “Subsidiarity Principle”, also sometimes referred to as the “Decentralization Theorem” (Oates, 1972, 1993, 1999).

An exception to the Subsidiarity Principle may occur in cases where the service has widespread spillovers, such as a public good whose benefits transcend localities. National defense is
a classic example here. As we discuss below, environmental quality may fall into this category as well, in cases where pollution crosses the borders of local jurisdictions or where a natural resource is widely valued (e.g., a U.N. heritage site). A second exception may be cases where services can be provided more efficiently at larger scales than the local jurisdiction. An example may be municipal solid waste: even large cities may benefit from sharing a single landfill rather than procuring their own individually, and even small towns and villages may benefit from sharing a single trash collection service. In such cases, if there are multiple levels of government, a middle level may be most suitable; alternatively, local governments may form their own compacts voluntarily.

These considerations suggest a “best practice” assignment of expenditure responsibilities across the different levels of government. Activities such as national defense, monetary policy, and income redistribution are appropriately assigned to the central government; activities like police and fire protection, trash collection, and local roads should be assigned to local governments. Of course, the actual practice on assignment of expenditure responsibilities differs somewhat from these “best practice” assignments. Nevertheless, despite substantial variation in expenditure assignments, the broad principles of assignment are generally upheld in most countries.

However, the argument for decentralization is weakened if the costs of local provision are higher due to the smaller scale of local government operations. Importantly, the argument for decentralization is also weakened if there are substantial spillovers from local government expenditures.

In this regard, consider “emissions control” or “pollution cleanup” as a government expenditure responsibility. Which level of government should be assigned its responsibility?
Put differently, which level of government should be assigned the responsibility for “environmental quality”? Environmental quality is clearly a good that has externality – and public good – aspects. However, there are different types of environmental quality, and the answer to the expenditure assignment question depends on the precise form that environmental quality actually takes. We discuss this issue in more detail below.

The Assignment of Tax Responsibilities

Once expenditure responsibilities have been assigned, tax instruments must also be assigned among the levels of government to provide adequate financing for required expenditures. Although there is much diversity in the fiscal structures of national and local governments, several “best practices” have emerged that provide a useful point of departure (Musgrave, 1983; McLure, 1994; Bird, 1999):

- Only the central government should impose progressive income taxes. Due to the potential mobility of factors, local government attempts to redistribute income by progressive income taxes will lead to the out-migration of more mobile, higher income individuals, thereby leaving more immobile, lower income individuals to bear the burden of the taxes. Income taxes are also thought to be effective countercyclical instruments, and macroeconomic goals are best pursued by national government policies.

- The central government should impose taxes on those tax bases that are distributed unequally across jurisdictions, and use the revenues from these taxes to equalize fiscal capacities across these areas.

- Local governments should rely predominately upon user charges and taxes on immobile tax bases; in particular, user charges should be used to finance goods that provide measurable benefits to identifiable individuals within a single jurisdiction, and taxes should be used to finance local services for which it is difficult to identify individual beneficiaries and to measure individual costs and benefits. The assignment of taxes should also meet the test of administrative feasibility.

- Local governments should avoid taxes on mobile tax bases, especially capital. As with progressive income taxes, the potential mobility of capital or other mobile factors of production will lead to out-migration if these factors are taxed at higher-than-average tax rates. By the same token, attempts to induce in-migration of mobile factors can lead to a so-called “race-to-the-bottom”, as local governments compete with each to attract and to hold these factors by extending tax breaks and other fiscal incentives.
Local governments should be assigned adequate sources of revenues consistent with their expenditure responsibilities. Local governments should have discretion over the rate of some taxes to promote accountability of local officials and to establish a link between services demanded and the cost of service provision. Locally assigned taxes should exhibit adequate revenue elasticity so that collections can grow with the demand of services over time.

Intergovernmental transfers should be used to finance those services that generate spillovers to nearby jurisdictions, since strictly local finance will lead to inefficient provision. Suppose, for example, that there are two local jurisdictions each of which provides an impure public good (e.g., pollution abatement) whose benefits spill over to the other jurisdiction. It can be shown that each locality should receive a subsidy (e.g., a “conditional”, “matching”, and “open-ended” grant) on its public good whose magnitude is equal to the marginal benefit of the externality, along the same lines as the earlier Pigouvian tax/subsidy (Oates, 1972; Alm, 1983; Gordon, 1983).

In practice, few countries rigidly follow these guidelines, although the broad pattern of tax assignment is often largely consistent with these prescriptions.

Indeed, around the world there are essentially two basic models of revenue assignment that attempt to satisfy these principles. In what might be called the Western or Anglo-Saxon model of “fiscally strong local governments” (e.g., the United States, Canada, Australia), local governments independently legislate and administer their own taxes, an approach that obviously gives local governments significant fiscal autonomy and adequacy. However, this model is probably not appropriate for many countries. Instead, in many other countries the model is one of “fiscally weak local governments” that do not generate significant levels of revenues from their own sources, that do not independently legislate and administer their own taxes, but that are often allowed to add a local tax onto the back of some existing central government tax. This approach is increasingly being used as part of decentralization reforms around the world, as discussed by Ahmad and Brosio (2006), Smoke, Gomez, and Peterson (2006), and Bird and Ebel, 2007). Also, see Bordignon and Ambrosanio (2006) for recent information on revenues and their assignment, focusing on trends in the European Union.
Of course, tax systems are designed to achieve multiple objectives. An obvious purpose is to raise the revenues necessary to finance government expenditures, and also to ensure that the growth in revenues is adequate to meet expenditure requirements. Another is to distribute the burden of taxation in a way that meets with a society’s notions of fairness and equity. Equity is typically defined in terms of “ability to pay”, such that those with equal ability should pay equal taxes (“horizontal equity”) and those with greater ability should pay greater taxes (“vertical equity”). Taxes can also be used to influence behavior of those who pay them; in choosing taxes, a common goal is to minimize the interference of taxes in the economic decisions of individuals and firms. Taxes should be simple, both to administer and to comply with, because a complicated tax system wastes the resources of tax administrators and taxpayers. The appropriate design of taxes requires balancing tradeoffs among these various goals.

However, the implications of locally generated externalities for the appropriate assignment – and design – of tax instruments have not been fully considered or analyzed. These issues are discussed later.

The Roles of Inter-jurisdictional Tax Competition and Globalization

Local tax systems in most all countries were originally designed for a world in which production and consumption were primarily of tangible goods, in which the sale and consumption of these goods generally occurred in the same location, and in which the factors of production used to make the goods were for the most part immobile. In such a world, taxation was a fairly straightforward exercise. Sales and excise taxes could be imposed on the tangible goods that were consumed, by the government in the jurisdiction in which consumption (or production) occurred. Similarly, income and property taxes could be imposed on factors where they lived and worked without fear that taxes would drive the factors elsewhere. In making these
tax decisions, a government in one jurisdiction had no need to consider how its actions would affect the governments in other jurisdictions because tax bases were largely immobile.

There is little doubt that, in principle, decentralization and other associated trends (especially competition among local governments and “globalization”, defined loosely as increased factor mobility across jurisdictions) changes things, and changes these decisions dramatically. Indeed, the world is very different today. In some nations, a trend toward fiscal decentralization has put more pressures on local tax systems, widening disparities across regions and increasing the importance of local taxes in the locational decisions of mobile factors. Ironically, the opposite trend of globalization has much the same effect. These forces have several main effects.

First, in a global economy, financial capital, firms, and even households are more mobile, making it harder for local – and even central – governments to raise revenues. For example, businesses have more flexibility in choosing where to locate because communication and transportation costs have been slashed. Further, some forms of production activity require little in the way of traditional capital and labor, so that physical location becomes less important. Labor, especially skilled labor, also becomes more mobile in this environment. Likewise, financial capital is able to flow quickly across local (indeed, state and national) boundaries.

Clearly, if factors of production can move easily from one location to another, then the ability of a local government to tax these factors is greatly diminished. A government that raises its tax rates above those of other jurisdictions risks losing its tax base to these areas. Particularly in the case of income from capital, there is much speculation that taxation will become increasingly problematic (Mintz, 1992). In fact, there is some empirical evidence (even if not yet
conclusive) that factors are responding to these types of tax considerations (Grubert, 1998; Hines, 1999).

Increased mobility is not limited to factors of production. Consumers are also able to plan their consumption according to tax considerations, and consumption does not necessarily occur in the jurisdiction in which a taxpayer resides. A jurisdiction that attempts to tax, say, gasoline more heavily than surrounding areas will find that consumers will purchase elsewhere. Similarly, individuals even in developing countries can now purchase many types of products over the internet and thereby avoid paying some (or even all) sales taxes. Additionally, there has been increased consumption of services and intangible goods (e.g., computer and information services) that are much more difficult to tax than tangible goods. The once-tight link between the location of sales and the location of consumption is now quite loose.

Second, and relatedly, the measurement, identification, and assignment of tax bases are now much more difficult. Consider a typical multi-jurisdictional business. The product that the firm makes may be designed in one or more jurisdictions; the firm may use inputs purchased in multiple jurisdictions; the product may be produced in several places and assembled in a still different location; and the final good may be sold in multiple locations. Because the business operates in multiple jurisdictions, the firm has considerable leeway to manipulate prices to minimize its tax liabilities. This latter problem is well known, but its severity has increased with the enormous expansion in the number of firms operating in multiple jurisdictions. It may be possible to overcome these types of issues by legislation, including international agreements (Tanzi, 1991, 1005), although these efforts remain uncertain. It should be noted that there is increasing empirical evidence that firms are making transfer pricing decisions to exploit these
types of opportunities (Swenson, 2001; Bartelsman and Beetsma, 2003; Clausing, 2003; Huizinga and Nicodeme, 2006).

Likewise, consider an individual whose income comes from multiple sources. A global income tax requires that income from these sources be aggregated. However, it is easy for an individual to hide, say, interest income from multiple areas. In the absence of information sharing across governments, the ability of a local government to identify incomes from other jurisdictions is quite limited.

Consider finally a consumer who can purchase goods and services in several different ways: from traditional local merchants or from company websites. In the former case, identification, measurement, and assignment of the tax base are straightforward. In the latter case, they are not. The application of, say, sales taxes in this new environment poses considerable problems for governments.

How will local governments respond to these various pressures in their tax choices? Most importantly, as emphasized throughout, the ability of any government to choose its tax policies independently of those in other jurisdictions is greatly curtailed. In the presence of mobile tax bases, a single government’s choice of tax policies will have effects beyond its own borders and will be affected by the actions of other jurisdictions. Accordingly, the analysis of tax choices by local governments must recognize that a local government will consider that its actions may elicit responses by other local governments as they respond in a strategic manner. These strategic interactions will have a number of effects.

The level of tax rates seems likely to decline. In particular, if tax bases can move easily from one jurisdiction to another, then they will flow from high-tax to low-tax areas. Owners of capital, skilled labor, and consumers will become increasingly sensitive to tax differentials in
their locational decisions. As a consequence, it is commonly argued that governments will face increasing pressures to compete with one another by reducing tax rates or by offering special tax incentives, in order to attract and to retain the various tax bases. For example, when a government reduces its tax rates on capital income, it thereby attracts capital flows from other jurisdictions, and in doing so the government benefits its own jurisdiction. However, the government’s action also imposes costs on the jurisdictions that lose factors of production, and it risks generating similar tax-cutting responses from those governments. With tax competition, there could well be a “race-to-the-bottom”, in which overall tax collections decline as local governments compete via low tax rates to attract or to retain their tax bases. To date, however, the evidence here is somewhat mixed and inconclusive; see Wilson (1999) and Wilson and Wildasin (2004) for comprehensive surveys. Also, if in fact there is a decline in the level of tax rates, then the level of government expenditures is also likely to decline.

The composition of local taxes could also change as a result of increased difficulty in taxing mobile tax bases. The overall tax burden from income taxes on mobile tax bases like capital and skilled labor will likely decline across local governments; tax rates on these factors should also flatten and converge. In contrast, taxes on immobile bases – unskilled labor, physical capital, and property – should increase. Of special note for the use of EIs, charges and fees for specific services should rise in importance because these tax bases are largely immobile. Local governments seem likely to turn more frequently to environmental or “green taxes”, as well as to “sin taxes” on alcohol and cigarettes and to lotteries, in attempts to replace lost revenues from mobile bases.

These compositional changes imply that local tax systems will likely become more regressive than at present. If taxes on capital and skilled labor decline, and if fees and charges,
sin taxes, income taxes on unskilled labor, and lotteries all increase, and if marginal income tax rates flatten, then local governments will find it quite difficult to maintain any progressivity in their tax systems. Together with an expected decline in overall revenues, the ability of local governments to redistribute income to lower income individuals will likely diminish.

The form of local sales taxes is also likely to change. Local (and other subnational) governments may well decide that a destination-based consumption tax that is collected by the federal government and distributed to them would be preferable to further erosion in their sales tax collections. Alternatively, they may agree among themselves to apply a uniform local sales tax. They may even radically reform the sales tax by moving toward a consumption-based, uniform-rate, destination-principle sales tax, as advocated by McLure (1997) and Fox and Murray (1997), among others.

These latter changes suggest more broadly that local governments may attempt greater harmonization (or at least coordination) of their tax systems, in an attempt to reduce the negative (fiscal) externalities that one government’s decisions impose upon other governments as well as to harmonize and coordinate environmental policies. Such harmonization implies that there should be some convergence in tax rates across local governments, and also in the definitions of tax bases. With harmonization, local autonomy in tax policy will obviously diminish (Tanzi, 1991, 1995, 2001). Central governments may effectively induce such harmonization through a system of intergovernmental transfers. If local governments cannot or will not provide adequate environmental protection, say because of competition for mobile capital or because they do not adequately account for inter-jurisdictional environmental spillovers, central governments can increase the level of protection via matching grants (Alm, 1983; Gordon, 1983).
Whether all these changes are good or bad is obviously difficult to determine. Most previous discussion has focused mainly on the negative fiscal externalities of tax competition (e.g., the race to the bottom), as captured especially in the work of Zodrow and Mieszkowski (1986), Wilson (1986), and Wildasin (1989). This work concludes that tax competition lowers welfare by reducing the level of tax rates (and so of public expenditures) below their efficient levels, as local governments attempt to compete for mobile capital by offering fiscal inducements (e.g., lower taxes) to businesses in a race to the bottom. This work also often concludes that tax competition lowers welfare by distorting the allocation of capital across jurisdictions.

Perhaps surprisingly, however, it can also be demonstrated that, even in the presence of fiscal competition between local governments, local government fiscal choices can sometimes achieve efficiency (Oates and Schwab, 1988; Wildasin, 2003). The crucial difference between these strikingly different conclusions depends largely upon the types of tax instruments that are available to local governments, the composition of the electorate, assumptions made about the presence (or the absence) of strategic local government behavior, and the time dimension (e.g., a static versus dynamic setting).

Further, it can be demonstrated that fiscal competition both limits and expands the choices that local governments can make, again with potentially positive impacts on efficiency. With greater factor and tax base mobility, local governments have more power to influence the locational decisions of firms, workers, and consumers. Those governments that succeed in these choices will be the ones that are better able to match taxes with expenditures, and so that are better able to give taxpayers the services (including environmental protection) that individuals wish for the taxes they pay, along the lines of Tiebout (1956) and his “voting with your feet” model. With greater factor and tax base mobility, local governments may also be forced to
reduce inefficient overspending; that is, government spending may be excessively large in the absence of tax competition, so that tax completion acts to reduce such inefficient spending. As argued by the “Leviathan” model of Brennan and Buchanan (1980), government spending tends to be excessive and inefficient. Constraining government spending via tax competition can therefore be welfare-enhancing, as demonstrated by Rauscher (1998), Edwards and Keen (1996), Gordon and Wilson (2003), and Keen and Kotsogiannis (2004), among others.

Overall, then, although much previous research has focused mainly on the negative fiscal externalities of tax competition, some recent work demonstrates the potentially positive effects of tax and, especially, of expenditure competition. It is these positive effects that deserve more emphasis. Again, see Wilson (1999) and Wilson and Wildasin (2004) for comprehensive surveys.

To illustrate these forces, consider a world in which all factors of production are completely mobile, there are no transportation or communication costs, and there is a single national market for all goods and services. It might seem that no government at any level would be able to impose taxes in such a hypothetical world because any taxes would lead to the immediate outflow of the tax base from the jurisdiction. Put differently, it might appear that the “vanishing taxpayer” would lead inexorably to the virtual disappearance of government, as suggested by Zodrow and Mieszkowski (1986), Wilson (1986), and Wildasin (1989).

Of course, these assumptions are extreme: complete mobility does not now, and will never fully, exist. Even so, this view is surely wrong. Individuals and firms value the goods and services that local governments provide, and they are willing to pay for them. As originally argued by Tiebout (1956), individuals will “vote with their feet” by moving to those jurisdictions in which governments provide services that residents value. Indeed, local governments will be
encouraged, even required, to make their communities as attractive as possible: by providing uncongested roads, a clean environment, pleasant parks, quality schools, safe neighborhoods, and the like, all with a tax burden that individuals deem responsible and appropriate. Local governments may also be forced to reduce inefficient overspending on public services, as suggested by Brennan and Buchanan (1980) and as demonstrated by Edwards and Keen (1996), Rauscher (1998), and Keen and Kotsogiannis (2004). (If individuals value redistribution, as many certainly do, then even programs for the poor would survive, albeit at smaller levels than currently.) Firms will not object because they too benefit from safe neighborhoods and quality infrastructure, as well as from the availability of workers who are attracted by these things. Moreover, this logic holds under the extreme assumption of mobility, so it should even more so under realistic levels of mobility.

In sum, then, even in an increasingly decentralized (and integrated) economy, local governments will still exist, they will still impose taxes, and they will still make expenditures. There is no question that these decisions will be circumscribed by the possibility of a “vanishing taxpayer”. However, the existence of such a taxpayer also creates opportunities, by giving local governments the potential to influence these locational decisions and by requiring local governments to become more efficient in their service delivery. Local governments whose prior performance has been poor will have little credibility in making policy decisions; the response to those governments with sound institutions will be quite different. There will therefore be pressures on all local governments to establish these institutions. Those governments that succeed in these choices will be the ones that are better able to match taxes with expenditures, or are better able to give taxpayers the services they desire for the taxes they wish to pay.
4. Designing Economic Instruments in a Decentralized Fiscal System

The basic analysis of EIs does not take into account the level of government addressing the problem, and it also does not take into account either the interactions among jurisdictions at the same level of government or the interactions across levels of a federalist system. By the same token, the standard analysis of the division of fiscal responsibilities does not account for the environmental consequences of those relationships. In this section, we draw on both literatures to consider the appropriate role for various levels of government when internalizing externalities, in different settings. Specifically, we apply the above insights on both EIs and fiscal federalism to the question of which level of government is best situated to address environmental problems. We begin with two considerations drawn from the literature on EIs, and we then turn to additional insights from the literature on fiscal federalism.

A first consideration is that, to work effectively, *EIs must be backed up by governmental monitoring and enforcement*. Central or local governments may differ in these abilities. Depending on how enforcement activities are funded (e.g., out of pollution penalties versus out of general revenues), the level of government with more resources may have an advantage. Likewise, the level of government that has the legal authority to enforce violations also may have an advantage. Finally, the level of government that is freer from corruption is better able to enforce EIs. This level may be the central government because it may be less beholden to any particular interest, or it may be local governments because they may be more accountable to local populations. Clearly, the factors discussed earlier have differing consequences, on a case-by-case basis, for which level of government is most appropriate to operate a given instrument or for which instrument is most appropriate for a given level of government.
A second consideration is the fact that market-based EIs have their greatest advantage when there is substantial cost heterogeneity. If all relevant agents are (nearly) identical, a one-size-fits-all policy would be perfectly appropriate. However, when agents differ in, say, their costs of pollution abatement, there will be efficiency gains from the flexibility that EIs provide in accommodating who abates pollution the most. By the same token, a cap-and-trade policy in particular will have its greatest advantage when there are more (and more different) firms involved in the trading. For this reason, if a country is a likely candidate for a cap-and-trade program, it will do better to implement it at a national level rather than to have multiple trading umbrellas within the country, without trade between them.

Our next two considerations come from the literature on environmental federalism, a natural extension of the fiscal federalism literature (Oates, 2002; Braden, Folmer, and Ulen, 1996; Dalmazzone, 2006). If emissions control (or environmental quality) is considered a government expenditure responsibility, then which level of government should be assigned its responsibility? As with the analysis of expenditure responsibilities, an important third consideration is that the appropriate level of government at which to address an environmental problem is determined by the geographic scope of the externality.

Following Oates (2002), we can classify pollution as being one of three types; see also Braden, Folmer, and Ulen (1996), Anderson and Hill (1997), Farber (1997), and Ogawa and Wildasin (2009) for relevant discussions. In a first type, the overall level of environmental quality for the nation as a whole depends upon the aggregate level of “perfectly mixing” waste emissions from all areas. The actual level of environmental quality in a particular local area may vary across localities, depending on, say, local weather conditions. Even so, however, it is the total amount of emissions from all localities that determines the exact level of environmental
quality in any local area. Examples include global climate change and the depletion of the ozone layer, where indeed aggregate worldwide emissions determine the environmental damage at a given location.

A second type of pollutant is one in which the amount of waste emissions in a given locality \( i \) depends only upon the waste emissions in that locality; that is, waste emissions in locality \( i \) impose costs on residents and business in locality \( i \), but there are no spillovers from the pollutants beyond the locality itself. An example of this second type of pollutant is the effects of local waste emissions on water quality in the locality, with no spillovers to other jurisdictions (i.e., where the water basin lies entirely within the jurisdiction). Another example is the collection and disposal of municipal solid waste.

In a third type of polluting activity, local waste emissions in locality \( i \) have harmful effects in that locality, and these waste emissions spill over to at least some immediately surrounding jurisdictions, without necessarily creating an “aggregate” externality. Similarly, pollutants from some other neighboring localities \( j \neq i \) spill over into jurisdiction \( i \) to create a level of environmental quality in \( i \) that depends upon emissions from \( i \) and from other surrounding jurisdictions. In practice, this third type of waste emissions, which lies between the two previous polar cases, seems likely to be the most common. Examples include acid rain and shared water resources.

In determining the appropriate level of government at which to address a pollution problem, a third consideration is therefore which of these three types of polluting activities is being considered. In the first type, where the overall level of environmental quality for the nation as a whole depends upon the aggregate level of waste emissions from all local areas, it is the national government that should be assigned the responsibility for the level of environmental
quality. Here, environmental quality has the characteristic of a pure public good at the national level. The central government can consider the full national effects of waste emissions, while local governments are likely to ignore inter-jurisdictional spillovers.

However, in the second case, it is the local government in each jurisdiction that should be assigned the responsibility for determining the efficient level of environmental quality in its jurisdiction. Because local waste emissions do not spill over to any other localities, the local government should be able to determine the efficient level of the purely local public good, environmental quality. In contrast, uniform national standards are likely to be inefficient because they will not allow local governments to adjust environmental quality either to the demands of local citizens or to local conditions that affect the cost of or optimal approach to pollution control. Dinan, Cropper, and Portney (1999) provide an example of municipal water quality in the U.S., in which federal regulations impose significant costs on some cities disproportionate to benefits.

The third case is the most difficult, where local waste emissions in locality i have harmful effects in that locality and these waste emissions spill over to some (though not all) immediately surrounding jurisdictions. Here, as Oates (2002) noted, there is a trade-off between relying on local governments, which tend to ignore spillovers, or on a central government, which tends to impose inefficient uniformity across local jurisdictions. Which choice is the lesser of the two evils will differ on a case-by-case basis. Garcia-Valiñas (2007) shows the importance of these trade-offs for water quality in Spain. Banzhaf and Chupp (2010) consider them for the case of air quality in the U.S. Banzhaf and Chupp (2010) also show that these trade-offs interact with the underlying production technologies in important ways. In particular, they point out that local jurisdictions tend to make the mistake of under-pricing pollution, whereas the central
government might at least be right “on average”. If marginal abatement costs are convex, this tendency implies that the local jurisdictions operate where costs are more elastic, causing larger deadweight losses, so on this dimension there may be a presumption in favor of more centralized policies.

As noted above, the central result in models of environmental federalism is that local jurisdictions can efficiently regulate local environmental quality (Oates and Schwab, 1988). However, economists have identified exceptions to this rule. These exceptions are the expenditure equivalent of the tax-side race to the bottom, in which local governments compete with one another to attract businesses, now by lowering environmental standards rather than by lowering taxes. Alternatively, local governments may engage in a race to the top to attract residents and export pollution onto other jurisdictions. *The potential for environmental races to the bottom or top, and other such strategic interactions among government*, is a fourth and final consideration for determining the appropriate level of government for dealing with pollution. The debate on an environmental race to the bottom first arose in the United States in the early 1970s, as the federal government began to displace the several states in environmental enforcement (Stein, 1971; Peltzman and Tideman, 1973). It continues to this day; for overviews and detailed discussions, see Revesz (1992, 1996), Wellisch (1995), Esty (1996), Oates and Schwab (1996), Anderson and Hill (1997), Engel (1997), Farber (1997), Oates (2002), Levinson (2003), Oates and Portney (2003), Kunce and Shogren (2005), and Dalmazzone (2006). It might appear that the possible existence of a race to the bottom in environmental quality is a factor generally in favor of a more centralized assignment of emissions control. However, the conclusions from this literature depend very much on the nature of both the environmental problem and the political behavior of local governments.
For example, Markusen, Morey, and Olewiler (1995) consider a world in which there are just two jurisdictions and one monopolistic and dirty firm. The jurisdictions set a level of environmental stringency for the factory; the firm decides where to locate. Because of transportation costs, consumers obtain the factory's output at lower costs if the factory is located in their jurisdiction; on the other hand, they are also then exposed to the local pollution. In this model, if environmental costs are lower than the gains to consumers from being near the factory, then the two jurisdictions will compete to host it, lowering environmental standards below the global optimum. However, if environmental costs are higher than the gain to consumers of lower prices, then jurisdictions will compete to raise their standards above the optimum level, trying to keep the pollution at a distance while still enjoying the imported product. Far from a race to the bottom, this case would be a “race to the top”.

Markusen, Morey, and Olewiler (1995) reach these conclusions because trade allows jurisdictions to reap the benefits of the firms’ production without the pollution and because the jurisdictions do not care about the monopolistic firm’s profits, so that the jurisdictions can attempt to capture those profits through the tax. Oates and Schwab (1988) provide a model with a very different set of assumptions; see Levinson (2003) for a detailed comparison and discussion. Oates and Schwab (1988) likewise assume that the pollution is only local, but the offsetting interests are those of workers rather than consumers. In their central model, all workers work in the dirty industry. In this case, there are no inter-jurisdictional externalities, and voters in each jurisdiction balance the costs of pollution against the gains in wages, a conclusion that would be strengthened by adding the Tiebout (1956) mechanism of households who “vote with their feet”. Jurisdictions that set environmental standards too stringently lose residents who
seek better jobs elsewhere. By the same token, jurisdictions that set standards too loosely also lose residents as they seek better amenities elsewhere.

However, Oates and Schwab (1988) also show that this conclusion can be weakened by a variety of auxiliary assumptions. First, to the extent that the pollution has cross-border effects, local jurisdictions would not be expected to take these into account. Second, jurisdictions may be forced to levy other (non-environmental) taxes on the industry to raise revenue; in this case, jurisdictions may compensate by lowering environmental standards. Alternatively, not all of the population may work for the dirty industry. In this case, the population will be divided into a group that cares only about the environment and a group that values both, and the optimal level of environmental stringency is a compromise (or balance) between the two positions, even though one group may win totally in the political solution. Finally, government officials may care more about the size of their budget than about the welfare of their citizens, an especially important consideration if the model is to be applied to a non-democratic society. In this case, officials may loosen environmental standards to maintain the tax base. Note, however, that this factor could work the other way if local jurisdictions employ a Pigouvian tax for the industry, rather than CAC regulations.

Others have shown how these conclusions also can depend on the specific EI employed. For example, Wellisch (1995) shows that, if local jurisdictions are using revenue-raising instruments, then they may internalize the value of emissions as a revenue source, not unlike the way emissions are passed through into wages in the model of Oates and Schwab (1988). In either case, local jurisdictions will efficiently balance the cost and benefits of emissions. On the other hand, if local jurisdictions are using command-and-control or other revenue-neutral
approaches, then they have an incentive to over-control pollution, as some of the burden falls on non-resident owners of capital.

If local governments cannot or will not provide appropriate environmental protection, say, because of competition for mobile capital or because they do not adequately account for inter-jurisdictional environmental spillovers, then central governments may need to play a larger role. However, this need not mean implementing environmental policy. Instead, just as with other public expenditures, the central government can increase the level of environmental protection via matching grants (Alm, 1983; Oates 1999). Moreover, Silva and Caplan (1997) and Nagase and Silva (2000) show that, by making redistribution conditional on local welfare (which in turn may be a function of all jurisdictions’ emissions), central governments can induce local governments to take transboundary externalities into account. This is a powerful insight, because it suggests that central governments have tools at their disposal that they can employ without knowing anything about abatement costs, the optimal way to abate in a local setting, or even the extent of transboundary externalities. So long as local governments know their externalities, the central government need only observe outcomes. Central governments can also play a role of coordinating the actions of more local governments. For example, the EU emissions trading scheme essentially links together a set of individual national-level cap-and-trade regimes (Convery and Redmond, 2007; Kruger, Oates, and Pizer, 2007). As noted previously, by expanding the extent of the market, these linkages have the advantage of expanding the gains from trade. However, in the first phase of the program, they also created incentives for member states to over-allocate emissions permits in the linked sectors, until the aggregate cap became essentially non-binding.
When evaluating these considerations, the question must be focus on which are the most relevant to any specific region and any specific environmental problem. Does the pollution cross jurisdictional boundaries? Do citizens gain materially from the health of the polluting industry? Is the industry mobile? Can the central government provide matching grants or other forms of incentives? These and other questions must be considered.

We must also look to the empirical evidence on the question. A particularly important question is the extent to which governments account for cross-border externalities in their policies. Sigman (2002) finds worse pollution on international rivers than on domestic rivers around the world, suggesting a tendency to ignore international spillovers. However, she finds less evidence of this effect on international rivers within the EU, where cooperation might be expected to be greater. Another approach is to look at inter-jurisdictional spillovers at the subnational level. Sigman (2007) has compared pollution readings in 47 nations. She finds some evidence that average pollution concentrations are higher in countries with federal systems and/or with more decentralized public expenditures, consistent with a race to the bottom. Interestingly, Sigman (2007) also finds that more federal nations have greater variability in pollution across locations, which is consistent with the idea that differences in populations’ tastes or in firms’ cost structures would be accommodated better by decentralized policies than by centralized ones.

The United States also provides an important opportunity to look at these questions because of its federal system. In U.S. environmental policy, the federal government typically mandates standards of environmental quality that must be met, but allows some degree of discretion on the part of individual states to implement and enforce these standards. Moreover, the Clean Water Act created a process whereby some states were “authorized” with more
discretion and others were not (Sigman, 2003). In that context, Sigman (2005) finds little overall effect of having more discretion on average pollution levels in a state’s rivers. However, she does find a small effect of state discretion on downstream states or on rivers that form a shared border; that is, where inter-jurisdictional spillovers are strongest, states that have the discretion to free ride seem to do so.

A related result comes from a study of the pattern of air and water emissions from the largest pollution sources in the U.S. Helland and Whitford (2002) find that both types of emissions are higher in border counties than in non-border counties. Since emissions near a border are more likely to spill over into neighboring jurisdictions, this finding suggests that states either are selectively enforcing regulations or are authorizing permits in a way that encourages such spillovers. Similarly, in a study of pollution from U.S. pulp and papers mills, which are subject to state-level regulations and enforcement, Gray and Shadbegian (2004) find that both air and water pollution emissions are lowest where their damages are likely to be highest (e.g., because of a more dense local population). This result is also consistent with the greater flexibility afforded by local rather than national policy.

Bluestone (2007) examines a similar issue, whether higher levels of “governmental fragmentation” (e.g., the number of local governments) in Metropolitan Statistical Areas (MSAs) leads to worse environmental outcomes. While the federal government sets environmental standards for air quality and the states determine how those goals are to be met, individual local jurisdictional policy can still have direct and indirect effects on outcomes. His empirical results suggest that local government fragmentation hinders MSAs from attaining of the ozone standard, a result consistent with the notion that local governments do not fully consider the effects of their policies on other neighboring jurisdictions.
Finally, an important test case arose for the U.S. in the 1980s, when states were given substantially more discretion to enforce air quality rules (and less assistance from the federal government) under Reagan’s "new federalism". List and Gerking (2000) and Millimet (2003) study the effect of this devolution of authority, and find that it had little effect on (and may if anything actually increased) expenditures on pollution abatement and air quality. Furthermore, Fredriksson and Millimet (2002) find no consistent pattern of a states’ environmental stringency on that of its neighbors.

The overall message in these studies appears to be that there is little evidence of a substantial race to the bottom from decentralized environmental policies. However, these studies also suggest that local jurisdictions tend to discount the importance of pollution on their neighbors. Thus, the evidence confirms the economist’s intuition that centralized policies will be especially important where these spillovers are strongest.

5. Conclusions

There are by now large, and largely separate, literatures both on the design of environmental instruments in a single unitary political system and on the design of fiscal institutions in a federal system with multiple and overlapping political jurisdictions. However, there has been less attention paid to the choice of environmental instruments in a federal system. We believe that there is much that can be learned by bringing these literatures together. We conclude by drawing several lessons from this synthesis. Our first perspective relates to the choice of which level of government should implement environmental instruments when such a choice is open. Our second perspective relates to the choice of which instrument might be most fitting at different levels of government.
If a fully functioning federal system offers some choice as to the level of government that should take action to reduce environmental pollution, our first and most important principle is the geographic scope of the externality. If the effects of waste fall within the same jurisdiction as the source, then local governments are probably best situated to address the externality, following the use of a standard and traditional EI. However, if the waste has significant transboundary effects, then the national government is better positioned to address it. Indeed, a locally generated externality that also contributes to a national externality requires a uniform central government response.

To this first principle, we add several mitigating factors. First, environmental policies must of course be enforced, so levels of government that have stronger police and judicial powers and that are freer from corruption are preferable. Second, even if the effects of pollution are local, if local sources of pollution are constituents of a national industry, then there may be efficiencies in imposing national rules. Third, and again even if the effects of pollution are local, local governments may be unable or unwilling to tackle the pollution problem because of fiscal competition. If it is a question of taxing the pollution or polluting industry, then local governments may fear the loss of jobs in that industry. Further, if it is a question of raising taxes on other sources in order to raise funds for (say) cleanup or for a revenue-providing instrument, then local governments may fear the loss of that alternative tax base. In these cases, central governments may prove more effective or may reduce these problems through revenue transfers. Again however, it is important to recognize the importance of a "race-to-the-bottom", which remains an open question.

Consider now the alternative perspective on the problem, or the choice of environmental instruments when the level of government is already determined. One lesson here is perhaps an
obvious one but one that must nevertheless be stated clearly. Namely, the instrument must be consistent with the fiscal authority of the government. In particular, a revenue-generating instrument must be tied to a treasury with the authority to levy and manage those taxes.

Our second principle is therefore that the instrument should be consistent with the fiscal needs of the level of government. If the jurisdiction is in need of revenues, a revenue-providing instrument will naturally be more appealing than a revenue-generating instrument.

Third, and similarly, the instrument should depend on the mobility of the polluting industry. Because polluters are naturally freer to move across local borders than to move across national borders, central government may have more ability to employ revenue-generating instruments on polluting industries, whereas local governments may face stronger political pressures to keep the industry. The local government may therefore prefer to use a revenue-neutral or even revenue-providing instrument. At the margin, these can provide the same incentive to polluters to reduce their pollution but can impose a lower total burden. However, local governments must also think about the mobility of the polluting industry relative to other tax bases. The funds for a revenue-providing instrument must come from somewhere. If the other sources of revenue are even more mobile than the dirty sources, it would not make sense to raise taxes on them further. Indeed it would make more sense to employ a revenue-raising instrument for pollution and reduce those other taxes.

A fourth principle is that ELs are likely to have their strongest advantage in the context of national policies. The reason for this is that much of their efficiency arises from the fact that they allow more abatement to be borne by polluters who have lower marginal abatement costs, and heterogeneity in these costs is likely to be greater the wider the geographic scope.
At the extreme extent of a pollution market, the arrival of global environmental problems such as climate change has posed well-known challenges for environmental policy-makers. However, as the world’s population continues to urbanize at the same time as governments are trending toward more decentralized authority, more local concerns must not be forgotten. Such pollution problems as local sanitation and water quality continue to be of tremendous human importance and raise policy challenges of their own. Given this range of environmental problems facing policy-makers, no single policy prescription could possibly be adequate for all problems. While the "right" policy choice will always be highly contextual, the literatures on environmental instruments and fiscal federalism provide important insights into the problem.

References


Bluestone, Peter (2007). *Governmental Fragmentation and Environmental Quality.* Unpublished doctoral dissertation, Andrew Young School of Policy Studies, Georgia State University. Atlanta, GA.


